Parametric and Bayesian non-parametric estimation of copulas

Dimitris Nicoloutsopoulos
University College London

A thesis submitted for the degree of
Doctor of Philosophy

2005

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

I would like to express my deep gratitude to my supervisor Dr. Kostas Skouras for his constant support, criticism and invaluable advice. Conversations with him have always been stimulating and enlightening, and are the sources of new ideas and creative thoughts. I am deeply indebted to Prof. Phil Dawid who was always available with his insightful comments and enthusiasm. He provided me with the greatest work environment at the department of Statistical Science, UCL and gave me good counselling throughout my research.

This work was supported by a studentship from the Greek State Scholarships Foundation (IKY).

Abstract

This thesis studies parametric and non-parametric methods of copula estimation with special focus on the Archimedean class of copulas.

The first part proposes an estimation procedure which is independent of the marginal distributions and performs well for one-parameter or two-parameter families of copulas, where traditional methods give questionable results especially for small sample sizes. In the second part we follow a Bayesian methodology and represent the copula density as a random piecewise constant function. Under the presence of some data, we set up a probability distribution over the copula density and utilize Markov Chain Monte Carlo techniques to explore that distribution. The methodology is extended to perform shape preserving estimation of a univariate convex and monotone function that characterizes the copula. The estimated first and second derivatives of the function of interest must satisfy the restrictions that the theory imposes.

All methods are illustrated with examples from simulated samples and a real-life dataset of the daily observations of the Dow-Jones and FTSE financial indices.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Investigating the relationship between a multidimensional probability distribution function and its lower dimensional margins is an interesting question which can be traced back to the works of Fréchet (1951), Féron (1956) and Dall'Aglio (1959), who studied bivariate and trivariate distributions with given marginals. The answer, for univariate margins, was given by Sklar (1959) with the introduction of a new class of distributions which he called *copulas*. In short, he proved that any multivariate distribution can be represented in terms of its underlying univariate margins by binding them together using a copula function. The fundamental theorem, which bears his name, states that any joint distribution function can be written as a function of the margins alone, and that function is a copula. If the margins are continuous, then the copula representation is unique. Conversely, if you have several margins and you plug them into a copula then you get a joint distribution that has those given margins and the dependence structure of the copula. The ability to extract from the joint distribution its margins and its dependence function is the cornerstone of most of the applications employing copulas. Typical examples are the investigation of dependence between random variables and the construction of multivariate families of distributions. As Fisher (1998) mentions, copulas are of interest to statisticians because they provide

- a framework to study scale-free measures of dependence and
- a starting point for constructing families of bivariate distributions.

An account of the events predated the introduction of copulas and the origination of the name, is given by Sklar himself in Sklar (1996) who cites: *While writing [...], I decided I needed a name for these functions. Knowing the word 'copula' as a grammatical term for a word or expression that links a subject and predicate, I felt that this would make an appropriate name for a function that links a multidimensional distribution to its one-dimensional margins and used it as such.* It should be noted that Hoeffding (1940) in his PhD thesis, '*Masstabinvariante Korrelationstheorie*' anticipated copulas which he called 'standardized distributions'. He investigated the Fréchet bounds and their interpretation in terms of functional dependence discussed desirable properties for a measure of dependence and used standardized distributions to construct such a measure.

For a long period after their introduction, copulas remained in a purely mathematical framework as they were used to develop the theory of probabilistic metric spaces. Menger (1942) proposed a probabilistic extension of the theory of metric spaces by replacing the distance $d(p, q)$ by a distribution function whose value $F_{pq}(x)$ is the probability that the distance between p and q is less than x. One of the first difficulties was to define the probabilistic equivalent of the triangular inequality. Menger proposed $F_{pq}(x + y) \geq T(F_{pr}(x) + F_{rq}(y))$, where T is a triangular norm or *t-norm*. Like a copula, a t-norm maps the unit square to the unit interval, and some copulas are t-norms while conversely, some t-norms are copulas.

Since the 1970s when the first statistical applications of copulas were presented, they have been rediscovered by several authors and referred to by different names such as 'dependence functions' by Deheuvels (1979) or 'uniform representations' by Kimeldorf and Sampson (1982). Subsequently it was discovered that copulas can be used to define non-parametric measures of dependence between random variables, and Schweizer and Wolff (1981) were the first to explicitly use copulas for the construction of such a measure. In the late 1990s copulas started being discussed in the actuarial literature, Tibiletti (1996), Frees and Valdez (1998) and a few years later, insurance mathematicians specialized in extreme value theory (EVT) made their marks in the field of finance by introducing copulas as a tool for financial risk management (Embrechts et al., 2002).

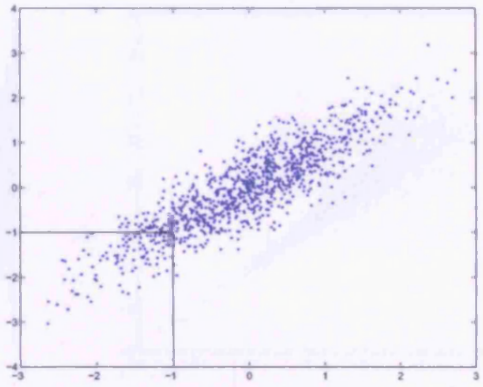


Figure 1.1: Scatter plot from a bivariate Normal with correlation 0.9. At low cut-offs correlation decreases. The bottom left rectangle includes approximately the 10% of the observations.

Financial institutions are required to estimate the Value at Risk (VaR) of all products in a portfolio every day. VaR is one measure of the potential change in the value of a portfolio with a given probability over a pre-defined horizon. For financial returns multivariate normality has two drawbacks. Firstly, the observed marginal distributions have heavier tails than those implied by the Normal distribution. In these cases, EVT provides a convenient framework for tail modelling. Secondly, it is believed that correlation increases in extreme cases (financial crises). More formally, Longin and Solnik (2001) have studied the asymptotic behaviour of the conditional correlation (i.e., the correlation conditionally on different values of one, or both, variables) of equity returns and using EVT have shown that these correlations deviate from what would be expected under multivariate normality. Specifically they define the extreme returns with respect to a threshold which determines the tail and derive the asymptotic distribution of the exceedances (i.e., returns above the threshold) as the threshold tends to the upper endpoint of the domain. In their words, *[...] asymptotic independence is often reached when the components of the return distribution are not independent. An important example is the multivariate normal distribution. [...]. For example, considering a bivariate normal process with standard mean and variance and a correlation of 0.8, the* (conditional) *correlation is equal to 0.48 for return exceedances one standard deviation away from the mean, 0.36 two standard deviations away from the mean, 0.24 three*

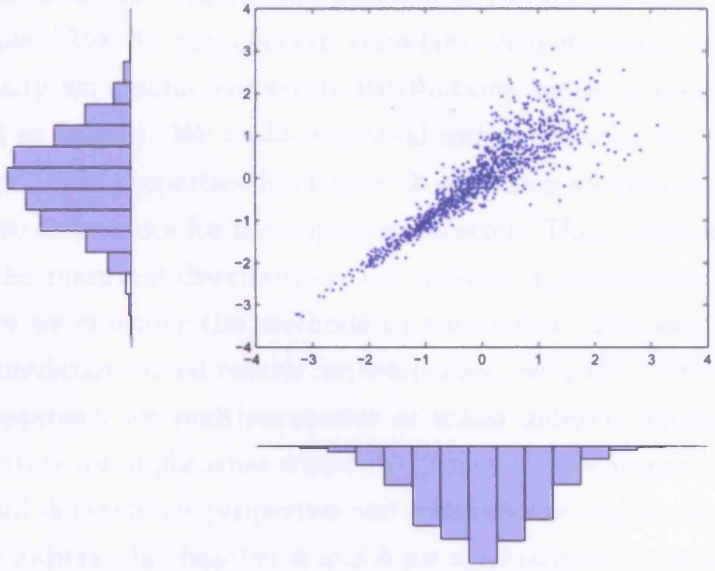


Figure 1.2: Scatter plot from a bivariate distribution with standard normal margins, correlation 0.9 and dependence structure that deviates from normality. Note that correlation increases at low cutoffs. The sample shown here has been drawn using the Clayton copula.

standard deviations away from the mean [...]. It goes to zero for extreme returns. We can graphically check the failure of the multivariate normal distribution to capture tail dependencies, in figure 1.1 where we have plotted 1000 observations from the bivariate normal distribution with standard normal margins and correlation 0.9. For low cut-off points as those shown in figure 1.1 and define the bottom left rectangle (consisting of small values), the correlation decreases.

The need for new ways to model correlation structure is possible to be met by copulas because they can generate distributions where correlation increases at the low or upper quantiles. A typical example is illustrated in figure 1.2 which shows a scatter plot from a bivariate distribution with standard Normal margins but completely different dependence structure from the Normal bivariate case. In this particular example, correlation increases at low cutoffs whereas it vanishes at high quantiles. Note that both samples shown in figures 1.1 and 1.2 have, apart from the same marginal distributions, the same correlation of 0.9.

In this thesis we investigate the problem of estimating the copula of a bivariate sample. We do not concern ourselves with marginal estimation and where necessary we assume empirical distributions for the margins. The thesis is structured as follows. We make a general introduction to the notion of copulas and their basic properties in chapter 2, and then we proceed to chapter 3 to present two estimators for the copula parameter. These estimators are independent of the marginal distributions but assume a parametric model for the copula, hence we consider the methods as parametric. For small sample size, we include simulation based results for comparison with the current margin-free estimation approach for multiparameter or scalar-indexed families of copulas. We also illustrate an application where we employ a two-parameter family with interesting tail dependence properties and estimate the copula of the returns of the financial indices. In chapters 4 and 5 we approach the estimation problem from a Bayesian perspective. More specifically, in chapter 4 we assume that the copula density is a random piecewise constant function with some prior distribution. Under the presence of some data we construct the posterior distribution and we draw an approximate sample in terms of a Markov Chain Monte Carlo (MCMC) methodology. The estimator for the theoretical copula density is the the sample mean of the MCMC sample. In chapter 5 we narrow down the scanning scheme by making the assumption that the data come from a specific generic class of families of copulas. This class has the nice property that it can be specified (up to a constant) from the generator, a univariate convex and decreasing function which becomes zero as it approaches the right endpoint of its domain. We assume that the second derivative of the generator is a stepwise function, while the first derivative and the generator itself are expressed as a linear and quadratic spline respectively. An MCMC approach is adopted to derive an approximate sample from the desirable distribution, the distribution of the second derivative given the presence of the data. Each observation is (deterministically) associated with its corresponding two anti-derivatives which are easily obtained. The estimator for the theoretical generator is then the sample mean of the appropriate MCMC sample. The first two derivatives are estimated similarly. The methodology we develop in this chapter falls within the area of estimation methods under shape restrictions, since the function under interest is convex and decreasing implying positivity and monotonicity of the second and

the first derivative respectively and these restrictions impose difficulties in the estimation procedure. The approach is also considered as non-parametric as the scanning concerns a functional space of infinite dimension and the estimation is dominated by the evidence contained in the observed sample. The methods presented in chapters 4 and 5 are illustrated on the same test datasets, one large of size $n = 2000$ and one smaller of $n = 500$. We also use the two estimators to estimate the copula of the real-life dataset from chapter 3 which consists of the daily returns of two financial indices. Chapter 6 concludes our investigations and includes quantitative comparative results for the two non-parametric Bayesian methods and the empirical estimator.

Chapter 2

An introduction to copulas.

2.1 Introduction.

In this chapter we introduce the notation and some basic properties for copulas (section 2.2). In section (2.3) we relate copulas to dependence measures and in section (2.4) we introduce an important class of copulas, the Archimedean class and discuss its most important properties. The construction of copulas is discussed in section (2.5) and the chapter concludes with a reference to investigations of Markov processes using copulas in section (2.6).

2.2 Basic definitions.

Definition 1. *Any multivariate distribution with uniform univariate margins on* $[0,1]$ *is a* ***copula.***

Mathematically, this definition is expressed as:

Definition 2. *An 2-dimensional copula is a function* $C : [0,1] \times [0,1] \rightarrow [0,1]$ *satisfying:*

- *(Boundary conditions)*

  $C(0, v_1) = C(u_1, 0) = 0$ *for all* $u_1, v_1 \in [0,1]$

  and

  $C(u_1, 1) = u_1$ *and* $C(1, v_1) = v_1$ *for all* $u_1, v_1 \in [0,1]$.

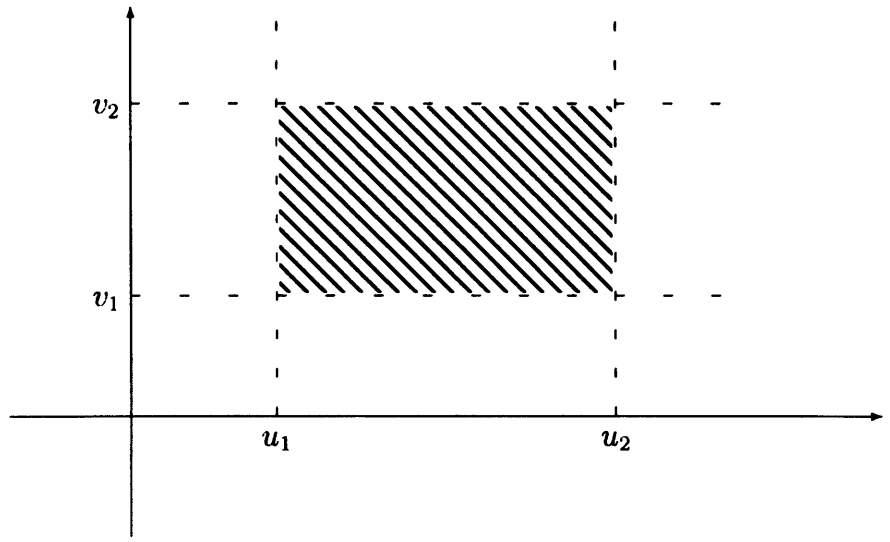


Figure 2.1: Copulas C assign a non-negative number on rectangles $[u_1, u_2] \times [v_1, v_2]$ from the unit square.

- *(Monotonicity condition)*

  $C(u_2, v_2) + C(u_1, v_1) - C(u_2, v_1) - C(u_1, v_2) \geq 0$ *for all* $u_1, v_2, u_1, v_2 \in [0, 1]$ *satisfying* $u_1 \leq u_2$ *and* $v_1 \leq v_2$

In this definition, the boundary conditions ensure that the margins are uniform distributions on the unit interval. The second condition assigns a positive number to each rectangle $[u_1, u_2] \times [v_1, v_2]$ from the the domain, (see figure (2.1)).

It follows now that for the rectangle $[u, 1] \times [v, 1]$:

$$1 - C(u, 1) - C(1, v) + C(u, v) \geq 0$$

and therefore $C(u, v) \geq u + v - 1$. Since $C(u, v)$ for any u, v, is always a non-negative number, we get:

$$C(u, v) \geq \max(u + v - 1, 0) \tag{2.1}$$

Furthermore, $C(u, v) \leq C(u, 1)$ and $C(u, v) \leq C(1, v)$, which yields:

$$C(u, v) \leq \min(u, v). \tag{2.2}$$

Inequality (2.2) combined with (2.1) shows that any copula $C(u, v)$ for any u, v from the domain, lies always between two surfaces:

$$C^-(u, v) \leq C(u, v) \leq C^+(u, v)$$

where $C^-(u,v) = \max(u+v-1, 0)$ and $C^+(u,v) \leq \min(u,v)$. Functions C^- and C^+ are called the *Fréchet-Hoeffding lower* and *upper* bounds respectively and they are themselves copulas. Another important copula which is commonly used is the product or independent copula: $C^{\perp\!\!\!\perp}(u,v) = \Pi(u,v) = uv$. Figure (2.2) shows perspective and contours plots for the Fréchet bounds and $C^{\perp\!\!\!\perp}$.

The following theorem from Nelsen (1999) establishes the continuity of copulas.

Theorem 2.2.1. *Let C be a copula. Then for every u_1, u_2, v_1, v_2 in the domain:*

$$|C(u_1, v_1) - C(u_2, v_2)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

Hence, C is absolutely continuous on its domain.

A further important property of copulas concerns the partial derivatives of C with respect to u and v and is summarized in the following theorem from Nelsen (1999).

Theorem 2.2.2. *Let C be a copula. For every $u \in [0,1]$ the partial derivative $\frac{\partial C}{\partial v}$ exists for almost all v. For such u and v:*

$$0 \leq \frac{\partial}{\partial v} C(u,v) \leq 1$$

For fixed v the function of u, $C_v(u) = \frac{\partial}{\partial v} C(u,v)$ is nondecreasing almost everywhere on the domain. The case for the partial derivative $\frac{\partial}{\partial u} C(u,v)$ is completely analogous.

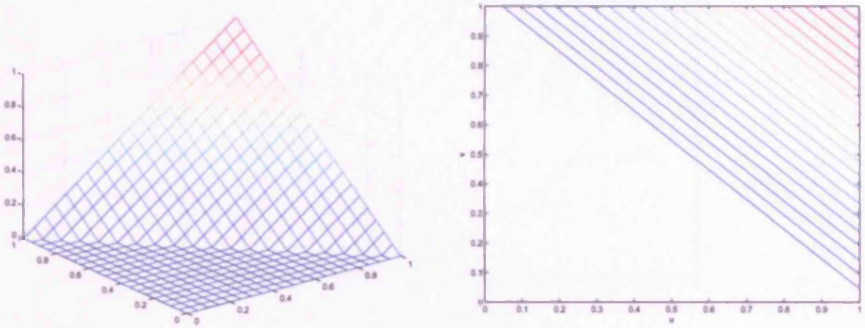
Example 1. *One of the copulas that we will encounter is the Gumbel - Hougaard copula:*

$$C(u,v) = \exp\Big(- [(-\log u)^\theta + (-\log u)^\theta]^{1/\theta} \Big)$$

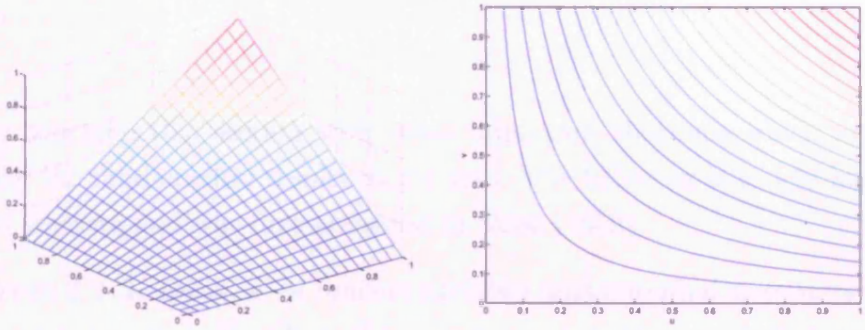
The partial derivative with respect to u is:

$$\frac{\partial}{\partial u} C(u,v) = C(u,v) \left[(-\log u)^\theta + (-\log u)^\theta\right]^{1/\theta - 1} \frac{(-\log v)^{\theta-1}}{v}$$

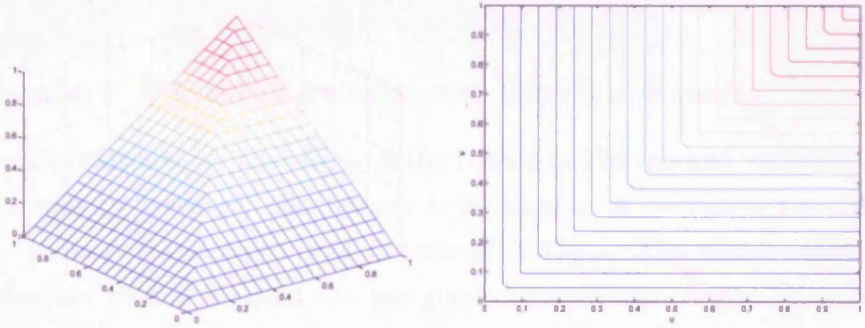
A plot of the function above is included in figure (2.3) for $\theta = 2$ and u ranging between $0.1, 0.2, \ldots, 0.9$.



(a) The lower Fréchet bound, $C^{-}(u,v) = \max(u+v-1, 0)$.



(b) The copula, $C^{\perp\!\!\!\perp}(u,v) = uv$.



(c) The upper Fréchet bound, $C^{+}(u,v) = \min(u,v)$.

Figure 2.2: Perspective and contours plots of the Fréchet upper and lower bounds. The middle plot shows the graph of the copula $C^{\perp\!\!\!\perp}(u,v) = uv$.

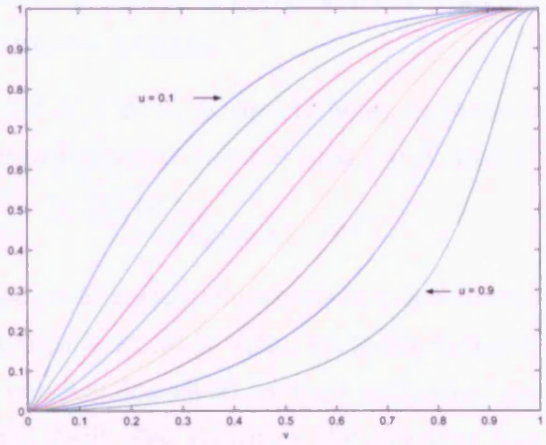


Figure 2.3: Plot of the partial derivative $\frac{\partial}{\partial u}C(u,v)$, for the Gumbel Hougaart copula .

We conclude this section with some important theorems which form the basis for the fundamental results in the area. The following is widely known as *Sklar's theorem* and was first appeared in Sklar (1959).

Theorem 2.2.3. *Let X, Y be random variables with marginal distributions F, G and joint distribution H. Then, there exists a copula $C_{X,Y}$ such that:*

$$H(x,y) = C_{X,Y}(F(x), G(y))$$

for all real x, y. If F and G are continuous then $C_{X,Y}$ is unique.

In such cases we say that $C_{X,Y}$ is the copula of the random variables X and Y. For brevity however, when there is no danger of confusion the subscript will be surpassed and write only C instead of $C_{X,Y}$. The name *copula* stems from the fact that it *couples* the marginal distributions to give the joint law. Note also that all marginal distributions to be encountered in this thesis will be continuous.

The following theorem concerns the characterization of independent random variables and the independent copula $C^{\perp\!\!\!\perp}$.

Theorem 2.2.4. *Let X, Y be random variables with continuous distribution functions F, G, joint distribution H and copula C. Then X and Y are independent if and only if $C = C^{\perp\!\!\!\perp}$.*

Another interesting property of copulas is that they remain invariant under strictly monotone transformations.

Theorem 2.2.5. *Let $X, Y F, G, H$ and C be as in theorem (2.2.4). If f and g are strictly increasing functions (on the range of X and Y respectively) then:*

$$C_{f(X),g(Y)} = C.$$

It becomes apparent now that the copula is a distribution on the ranks rather on the actual observations x and y. Also, since dependence between the random variables X and Y can not described by their individual margins F, G, it follows that it is the copula itself that captures the essential features of dependence. For this reason it is commonly said that the dependence structure is described by the copula. To underline this property the name 'dependence function' has been encountered with reference to copulas, (Deheuvels (1978), Kimeldorf and Sampson (1982)). Furthermore, since the copula does not change under strictly increasing transformations of the margins, it follows from theorem (2.2.3) that it is the copula that describes all these properties from the joint distribution which are invariant under such transformations. This property was exploited by Schweizer and Wolff (1981) to investigate rank statistics and construct non-parametric measures of dependence. This topic is discussed further in section (2.3).

An immediate consequence of theorem (2.2.3) is that if $F^{(-1)}, G^{(-1)}$ are the quasi-inverses of F and G respectively then for every u, v in the domain of C

$$C(u, v) = H(F^{(-1)}(u), G^{(-1)}(v)). \tag{2.3}$$

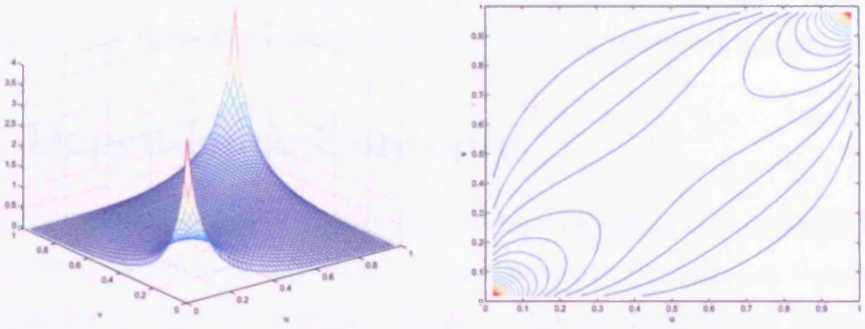
where the inverse is defined as : $F^{(-1)}(u) = \inf\{x : F(x) \geq u\}$.

One can now construct non-standard bivariate distributions using some copula C and employing margins F, G at will. We can also extract the dependence structure from a joint distribution by plugging in the inverses of the margins as (2.3) implies. For example, the function defined as

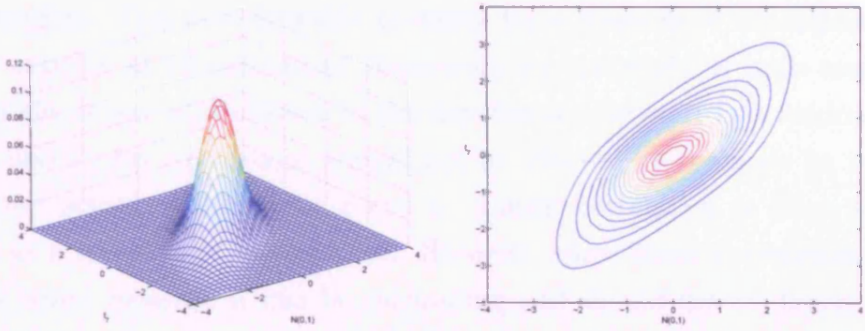
$$C(u, v) = N_r(\Phi^{-1}(u), \Phi^{-1}(v)) \tag{2.4}$$

which is explicitly written as:

$$C(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-r^2}} \exp\left(\frac{1}{2(1-r^2)}(s^2 - 2rst + t^2)\right) \, ds \, dt$$



(a) Perspective and contour plots of the normal copula density with $r = 0.7$.



(b) Perspective and contour plots of a distribution density with copula as above and standard normal and student's t (df = 7) margins.

Figure 2.4: Densities of two distributions. The upper half corresponds to the standard normal bivariate copula. This model combined with Student's t-margins with 7 d.f and standard normal margins, yields the density shown in the lower half.

expresses the dependence structure contained in a bivariate standard normal distribution with correlation r; hence the name, *bivariate standard normal copula*, Joe (1997). A plot of the density of the copula in (2.4) with correlation coefficient $r = 0.7$ is included in figure (2.4(a)). It is straightforward to construct other families of distributions by changing the margins at will. For instance, we can use the dependence structure in (2.4) and arbitrary margins. An example is illustrated in figure 2.4(b) where the plotted bivariate density has as margins a student's t_7 and a standard normal. Although this distribution has the dependence structure of the normal distribution, the difference with the standard bivariate normal model is apparent. In general, any distribution of the form

$$H(x, y) = C(\Phi(x), \Phi(y))$$

where C is not the normal copula of (2.4) expresses a bivariate distribution which is not the bivariate standard normal but has standard normal margins.

2.3 Dependence Concepts.

Since the dependence structure among random variables is represented by the distribution C, a natural way to study and measure dependence between random variables is through copulas. Many of the properties and measures associated with copulas, are invariant under strictly increasing transformations of the margins. This is a desirable property for a measure of dependence and suggests copulas as an convenient framework for the study of such measures. For example, if one is interested in the dependence between two random variables X and Y, his conclusions should not be affected by the scale he is using to take the actual observations x and y. Linear correlation is often used in practice as a measure of dependence. However, since linear correlation is not a copula based measure it can be misleading and should not be taken as the canonical dependence measure. A nice discussion on zero correlation versus dependence and relevant misinterpretations is presented in Drouet-Mari and Kotz (2001). The issue of having the correlation vanished even if there is dependence is discussed in most standard textbooks in probability or statistics; Casella and Berger (2002), Feller (1968), Feller (1971) to mention some. For additional discussion of the advantages of copula-based measures of dependence over linear correlation the reader may refer to Genest and Plante (2003) or to the survey by Embrechts et al. (2002). In Drouet-Mari and Kotz (2001) you may also find a very good coverage of a variety of measures of association as well as dependence ordering issues used by researches in the area.

We start with a short presentation of the linear correlation and then we continue with a discussion of some copula based measures of dependence.

2.3.1 Linear Correlation

Definition 3. *Let X and Y be two real random variables with finite variances. The linear correlation coefficient between X and Y is:*

$$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} \tag{2.5}$$

where $Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ *is the covariance between* X *and* Y, *and* $Var(X), Var(Y)$ *denotes the variances of* X *and* Y *respectively.*

Linear correlation is a measure of linear dependence. In the case of perfect linear dependence, i.e., $Y = \alpha X + \beta$ for α a nonzero real and β a real number we have: $\rho(X,Y) = sign(\alpha)$. In any other case $-1 \leq \rho(X,Y) \leq 1$. Furthermore, linear correlation has the property that:

$$r(\alpha X + \beta, \gamma Y + \delta) = sign(\alpha\gamma) r(X,Y)$$

for $\alpha, \beta, \gamma, \delta$ real and α, β nonzero. Clearly, for positive α and γ,

$$r(\alpha X + \beta, \gamma Y + \delta) = r(X,Y),$$

therefore, linear correlation is independent of the origins and units of measurements, that is, it remains invariant under strictly increasing *linear* transformations.

Linear correlation is easily manipulated under linear operations. Let $A, B \in \mathbb{R}^m$ and let $\boldsymbol{X}, \boldsymbol{Y}$ be random n-dimensional vectors. Then

$$Cov(A\boldsymbol{X} + \alpha, B\boldsymbol{Y} + \beta) = A\, Cov(\boldsymbol{X}, \boldsymbol{Y})\, B'$$

Following Lehmann (1966), the definition in (2.5) can be rewritten as:

$$r(X,Y) = \frac{1}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \int_{\infty}^{-\infty} \int_{\infty}^{-\infty} [H(x,y) - F(x)G(y)]\, \mathrm{d}x\, \mathrm{d}y$$

where F, G and H, the marginals of X, Y and their respectively. The probability transform (Whitt (1976)) that substitutes $u = F(x)$ and $v = G(y)$ yields:

$$r(X,Y) = \frac{1}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \int_0^1 \int_0^1 [C(u,v) - uv]\, \mathrm{d}F^{-1}(u)\, \mathrm{d}G^{-1}(v) \qquad (2.6)$$

where $C(u,v) = H(F^{-1}(u), G^{-1}(v))$.

Clearly, the last expression involves apart from the copula C the variances $Var(X), Var(Y)$ and therefore, linear correlation r is not invariant under strictly increasing transformations. In the following section we present the most important measures of dependence that are based on the ranks and as such they are independent of strictly increasing transformations.

Note that the various measures that have been used by researchers at times to summarize the association between random variables can be categorized into two broad classes; measures of dependence and measures concordance. To define the two groups, a set of desirable properties have been proposed.

The following is a set of desiderata for a measure of dependence δ adapted from Rényi (1959) and Schweizer and Wolff (1981).

(a) δ is defined for every pair (X, Y) of continuous random variables.

(b) $0 \leq \delta(X, Y) \leq 1$.

(c) $\delta(X, Y) = \delta(Y, X)$.

(d) $\delta = 0$ if and only if X and Y are independent.

(e) $\delta(X, Y) = 1$ if and only if each of X, Y is almost surely a strictly monotone function of the other.

(f) If f, g are almost surely strictly monotone functions on Range(X) and Range(Y), respectively, then $\delta(f(X), g(Y)) = \delta(X, Y)$

(g) If $(X_1, Y_1), (X_2, Y_2), \ldots$ is a sequence of continuous random vectors that converges weakly to a pair (X, Y), then $\delta(X_n, Y_n) \to \delta(X, Y)$ as $n \to \infty$.

The idea behind Schweizer and Wolff (1981) is that since the copula $C^{\perp\!\!\!\perp}(u, v) = uv$ characterizes independence, then for any two random variables X, Y any properly normalized distance between their copula C and $C^{\perp\!\!\!\perp}$ defines a measure of dependence between X and Y. Indeed, in the same paper the authors study the L_1, L_2 and L_∞ distances and show that the measures:

$$\sigma(X, Y) = 12 \int_0^1 \int_0^1 |C(u, v) - uv| \, du \, dv$$

$$\gamma(X, Y) = \left(90 \int_0^1 \int_0^1 \left(C(u, v) - uv\right)^2 du \, dv\right)^{1/2}$$

$$\kappa(X, Y) = 4 \sup_{u,v \in [0,1]} |C(u, v) - uv|$$

fulfill the conditions above and hence define a dependence measure. Note that the statistic $\gamma(X, Y)$, without the normalizing constant, was first introduced by

Blum et al. (1961) but Schweizer and Wolff (1981) were the first to put it under the more general copula-based framework.

While a measure of dependence summarizes the degree of the relationship by assigning a number between zero and one, with extremes at mutual independence and (monotone) dependence, it does not however give any information about the pattern of the relationship i.e., positive or negative association. The following set of rules defines a class of association measures which encompass the most popular rank-based measures used in literature.

According to Scarsini (1984), any measure of concordance κ should satisfy:

(a) κ is defined for every pair (X, Y) of continuous random variables.

(b) $-1 \leq \kappa(X, Y) \leq 1$, $\kappa(X, X) = 1$ and $\kappa(X, -X) = -1$.

(c) $\kappa(X, Y) = \kappa(Y, X)$.

(d) If X and Y are independent, then $k(X, Y) = 0$.

(e) $\kappa(-X, Y) = \kappa(X, -Y) = -\kappa(X, Y)$.

(f) If $(X, Y) \prec (X', Y')$ in the positive quadrant dependence ordering [1], then $\kappa(X, Y) \leq \kappa(X', Y')$.

(g) If $(X_1, Y_1), (X_2, Y_2), \ldots$ is a sequence of continuous random vectors that converges weakly to a pair (X, Y). then $\kappa(X_n, Y_n) \to \kappa(X, Y)$ as $n \to \infty$.

As a consequence of the definition, Nelsen (1999) states the following theorem:

Theorem 2.3.1. *Let κ be a measure of concordance for continuous random variables X and Y.*

1. *If Y is an almost surely increasing function of X, then $\kappa(X, Y) = 1$.*
2. *If Y is an almost surely decreasing function of X, then $\kappa(X, Y) = -1$.*
3. *If f, g are almost surely strictly monotone functions on the range of X and Y respectively, then $\kappa(f(X), g(Y)) = \kappa(X, Y)$.*

[1] Let (X, Y) be a bivariate random vector with joint distribution H. H, or (X, Y), is positive quadrant dependent (PQD) if $\mathbb{P}[X \geq x, Y \geq y] \geq \mathbb{P}(X \geq x)\mathbb{P}(Y \geq y), \forall x, y$. If (X', Y') is another random vector with joint cdf H', then H' is more PQD than H, denoted by $H \prec H'$, (or $(X, Y) \prec (X', Y')$) if $H(x, y) < H'(x, y)$ for all (x, y).

2.3.2 Kendall's tau.

Two widely used measures of concordance are Kendall's τ and Spearman's ρ. Before proceeding with the presentation of these measures, we clarify the term 'concordance'. Let (x_i, y_i) and (x_j, y_j) be two observations from a random vector (X, Y) of continuous random variables. We say that (x_i, y_i) and (x_j, y_j) are concordant if $x_i < x_j$ and $y_i < y_j$, or if $x_i > x_j$ and $y_i > y_j$. Similarly, we say that (x_i, y_i) and (x_j, y_j) are disconcordant if $x_i < x_j$ and $y_i > y_j$, or if $x_i > x_j$ and $y_i < y_j$. Alternatively, we can write that (x_i, y_i) and (x_j, y_j) are concordant if $(x_i - x_j)(y_i - y_j) > 0$ and disconcordant if $(x_i - x_j)(y_i - y_j) < 0$. Regarding the population, we can say that two independent pairs of variables (X, Y) and (X', Y'), with the same distribution, are called concordant if

$$\mathbb{P}[(X - X')(Y - Y') > 0] = 1$$

and disconcordant otherwise.

The Kendall's index τ is defined as:

$$\begin{aligned} \tau &= \mathbb{P}[(X - X')(Y - Y') \geq 0] - \mathbb{P}[(X - X')(Y - Y') \leq 0] \\ &= 2\,\mathbb{P}[(X - X')(Y - Y') \geq 0] - 1. \end{aligned} \tag{2.7}$$

Now, since

$$\begin{aligned} \mathbb{P}[(X - X')(Y - Y') \geq 0] &= \mathbb{P}[(X \geq X')(Y \geq Y')] + \mathbb{P}[(X \leq X')(Y \leq Y')] \\ &= 2\,\mathbb{P}[(X \geq X')(Y \geq Y')], \end{aligned}$$

equation (2.7) yields:

$$\begin{aligned} \tau &= 4\,\mathbb{P}[(X \geq X')(Y \geq Y')] - 1 \\ &= 4 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathbb{P}(X' \leq x, Y' \leq y)\, \mathrm{d}H(x, y) \end{aligned}$$

and by substituting $u = F(x), v = G(y)$ and $H(x, y) = C(F(x), G(y))$,

$$\tau = 4 \int_0^1 \int_0^1 C(u, v)\, \mathrm{d}C(u, v) \tag{2.8}$$

Given a sample $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$ we can estimate τ by:

$$\hat{\tau} = \frac{\#(\text{concordant pairs}) - \#(\text{disconcordant pairs})}{\binom{n}{2}}$$

2.3.3 Spearman's rho.

Spearman's index, ρ is defined by

$$\rho(X,Y) = 12 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \big(H(x,y) - F(x)G(y)\big)\, \mathrm{d}F(x)\, \mathrm{d}G(y),$$

or equivalently,

$$\rho(X,Y) = 12 \int_0^1 \int_0^1 (C(u,v) - uv)\, \mathrm{d}u\, \mathrm{d}v \tag{2.9}$$

where $F(x) = u, G(y) = v$ and $H(x,y) = C(F(x), G(y))$, Schweizer and Wolff (1981). Note that ρ is simply the normalized signed difference between two surfaces; the copula C of the random variables X, Y and the independence copula $C^{\perp\!\!\!\perp}(u,v) = uv$. Expanding (2.9) we get:

$$\begin{aligned}\rho(X,Y) &= 12 \int_0^1 \int_0^1 uv\, \mathrm{d}C(u,v) - 3 \\ &= 12\,\mathbb{E}(UV) - 3\end{aligned}$$

and since U, V are random variables uniformly distributed on $(0,1)$ it follows that $\mathbb{E}(U) = \frac{1}{2}$ and $\mathrm{Var}(U) = \frac{1}{12}$ and we can rewrite the last expression as:

$$\rho(X,Y) = \frac{\mathbb{E}(UV) - \frac{1}{4}}{\frac{1}{12}}. \tag{2.10}$$

Hence,

$$\rho(X,Y) = \frac{\mathbb{E}(UV) - \mathbb{E}(U)\mathbb{E}(V)}{\sqrt{\mathrm{Var}(U)\mathrm{Var}(V)}}$$

and therefore Spearman's ρ can be considered as the correlation coefficient for the random variables $U = F(X)$ and $V = G(Y)$.

A more thorough investigation of the rank-based measures can be found in Kruskal (1958) where in his words he *'discuses the probabilistic and operational interpretation of several well known measures of association particularly those that are ordinally invariant'*. The term *'ordinally invariant'* means that the measure is invariant under monotone transformations of the coordinates. We conclude this section with a simple example demonstrating that correlation coefficient does not remain unchanged by such transformations.

Example 2. *Let X, Y be exponential distributions with mean equal to one and copula C be a member of the Farlie-Gumbel-Morgenstern:*

$$C(u,v) = uv + \theta uv(1-u)(1-v),$$

for some $-1 \leq \theta \leq 1$. The joint distribution of X, Y is given by

$$H(x,y) = C(1-e^{-x}, 1-e^{-y}),$$

the correlation coefficient is

$$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \mathbb{E}(XY) - 1$$

and

$$\mathbb{E}(XY) = \int_0^\infty \int_0^\infty xy \; \mathrm{d}H(x,y) = 1 + \frac{\theta}{4}.$$

Therefore, $r(X,Y) = \theta/4$.

Suppose now that one transforms the random variables X, Y as:

$$U = F(X) \quad \textit{and} \quad V = G(Y),$$

then the correlation coefficient of U and V is

$$\begin{aligned} r(U,V) &= 12 \int_0^1 \int_0^1 C(u,v) \; \mathrm{d}u \; \mathrm{d}v - 3 \\ &= 12 \int_0^1 \int_0^1 (uv + \theta uv(1-u)(1-v)) \; \mathrm{d}u \; \mathrm{d}v - 3 \\ &= \frac{\theta}{3}. \end{aligned}$$

and hence, r can not be considered as a measure of concordance. This is not the case for either Kendall's τ or Spearman's ρ and the interested reader may find in Nelsen (1999) that both τ and ρ satisfy the properties of the definition for the measure of concordance.

2.3.4 Tail dependence.

The concept of tail dependence describes the amount of dependence in the lower-left-quadrant tail or upper-right-quadrant tail of a bivariate distribution. Distributions possessing tail dependence are able to incorporate dependencies

of extremal events. Furthermore, tail dependence is an expression of the copula C of the two random variables and as such, remains invariant under strictly increasing transformations of X and Y. The following definition has been adapted from Joe (1997).

Definition 4. *Let X, Y be continuous random variables with distributions functions F, G respectively. The coefficient of upper tail dependence of X and Y is:*

$$\lambda_U = \lim_{u \to 1} \mathbb{P}(X > F^{-1}(u) \,|\, Y > G^{-1}(u))$$

provided that the limit exists. If $\lambda_U \in (0, 1]$ then the joint distribution of X, Y has upper tail dependence, otherwise if $\lambda = 0$, no upper tail dependence exists.

If X and Y have copula C, $\mathbb{P}(X \leq x, Y \leq y) = C(F(x), G(y))$ and if we denote $\bar{C}(u, u) = \mathbb{P}(U > u, V > u) = 1 - 2u + C(u, u)$, then the coefficient λ_U can be rewritten as:

$$\lambda_U = \lim_{u \to 1} \frac{\bar{C}(u, u)}{1 - u} \tag{2.11}$$

which, as an expression of the copula, remains invariant to increasing transformations.

Alternatively, (2.11) is equal to:

$$\lambda_U = -\lim_{u \to 1} \frac{\mathrm{d}\bar{C}(u, u)}{\mathrm{d}u} = 2 - \lim_{u \to 1} \frac{\mathrm{d}}{\mathrm{d}u} C(u, u). \tag{2.12}$$

Because,

$$\mathbb{P}(U \leq u \,|\, V = v) = \frac{\partial}{\partial v} C(u, v),$$

we get from (2.12) that

$$\lambda_U = 2 - 2 \lim_{u \to 1} \mathbb{P}(U \leq u \,|\, V = u),$$

which finally yields,

$$\lambda_U = 2 \lim_{u \to 1} \mathbb{P}(U > u \,|\, V = u).$$

Example 3. *Let X, Y two real random variables jointly distributed according to the bivariate standard normal distribution with correlation coefficient r. Then:*

$$\lim_{u \to 1} \mathbb{P}(U > u \,|\, V = u) = \lim_{x \to \infty} \mathbb{P}(X > x \,|\, Y = x)$$

and since $X \mid Y = y \sim N(ry, 1 - r^2)$, *it follows that:*

$$\lambda_U = 2 \lim_{x \to \infty} \bar{\Phi}\left(\frac{x - rx}{\sqrt{1 - r^2}}\right) = 0$$

and therefore, the normal copula (for $r \neq 1$*) does not exhibit upper tail dependence .*

Example 4. *Let* C *be a bivariate copula from the Gumbel - Hougaard copula:*

$$C(u, v) = \exp\Big(- [(-\log u)^\theta + (-\log u)^\theta]^{1/\theta}\Big)$$

for $\theta \geq 1$.

Then

$$\frac{\bar{C}(u, u)}{1 - u} = \frac{1 - 2u + \exp(2^{1/\theta} \log u)}{1 - u} = \frac{1 - 2u + u^{2^{1/\theta}}}{1 - u}.$$

Employing L'Hôspital's rule we can show that the limit:

$$\lim_{u \to 1} \frac{\bar{C}(u, u)}{1 - u} = 2 - 2^{1/\theta},$$

and hence, the Gumbel-Hougaard copula has upper tail dependence for $\theta > 1$. *Note that the case where* $\theta = 1$ *corresponds to the independence copula,* $C^{\perp\!\!\!\perp}$.

The concept of lower tail dependence is analogous to upper tail dependence. Naturally, if the limit

$$\lim_{u \to 0} \frac{C(u, u)}{u} = \lambda_L$$

exists, then C has lower tail dependence if $\lambda_L \in (0, 1]$. If $\lambda_L = 0$ then C does not have lower tail dependence.

Example 5. *Consider the Clayton copula, defined by:*

$$C(u, v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}, \quad \theta > 0.$$

Then, the lower tail dependence is:

$$\lambda_L = \lim_{u \to 0} \frac{C(u, u)}{u} = \lim_{u \to 0} \frac{(2u^{-\theta} - 1)^{-1/\theta}}{u}.$$

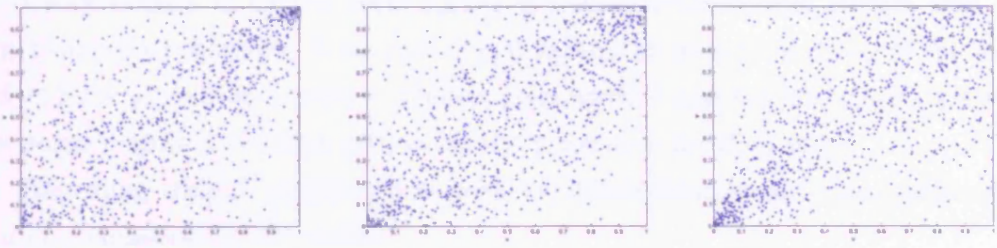
Elementary calculations show that

$$\lambda_L = \lim_{u \to 0} (2 - u^\theta)^{-1/\theta},$$

and finally, the lower tail dependence coefficient for the Clayton copula is:

$$\lambda_L = 2^{-1/\theta}. \tag{2.13}$$

Figure (2.5) illustrates three samples from three copulas, each one with different tail dependence characteristics.



(a) Gumbel copula. (b) Normal copula. (c) Clayton copula.

Figure 2.5: Samples of size 1000 from three copulas. The Gumbel-Hougaard copula has upper tail dependence only, and the Clayton copula, lower tail dependence only. The normal copula does not have either lower or upper tail dependence. All samples give estimated values for Kendall's τ close to 0.5.

2.4 Archimedean family.

In the previous section we encountered the copula

$$C(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}, \quad \theta > 0.$$

This can be expressed as

$$\left(C(u,v)\right)^{-\theta} - 1 = u^{-\theta} - 1 + v^{-\theta} - 1, \tag{2.14}$$

and therefore, it satisfies the relationship:

$$\phi(C(u,v)) = \phi(u) + \phi(v), \tag{2.15}$$

for

$$\phi(t) = t^{-\theta} - 1, \quad \theta > 0. \tag{2.16}$$

The relation (2.15) characterizes an important class of copulas widely known as Archimedean copulas. In this section we present the most important results in relation with this class. We start with the following theorem from Schweizer and Sklar (1983).

Theorem 2.4.1. *Let ϕ be a continuous, strictly decreasing function from $[0,1]$ to $[0,\infty]$ such that $\phi(1) = 0$ and let ϕ^{-1} be the inverse of ϕ. Then, the function C from $[0,1] \times [0,1]$ to $[0,1]$ given by*

$$C(u,v) = \phi^{-1}(\phi(u) + \phi(v)) \tag{2.17}$$

is a copula if and only if ϕ is convex.

Every copula C in the form (2.17) is called *Archimedean copula* and the function ϕ its *generator*. Note that C is invariant to transformations of the form $k\,\phi(t)$ for k a positive constant. Hence, generators ϕ and $k\,\phi$ are associated with the same copula C. When $\phi(0) = \infty$ the generator is called strict.

Assuming that ϕ'' exists and is positive, differentiation of C with respect to u and v equation (2.15) yields:

$$\phi'(C)\frac{\partial C}{\partial u} = \phi'(u)$$

and

$$\phi''(C)\frac{\partial C}{\partial u}\frac{\partial C}{\partial v} + \phi'(C)\frac{\partial^2 C}{\partial u \partial v} = 0.$$

Hence,

$$c(u,v) = \frac{\partial^2 C}{\partial u \partial v} = -\frac{\phi''(C)\phi'(u)\phi'(v)}{[\phi'(C)]^3}$$

The density $c(u,v)$ of C, is defined on all (u,v) such as $\phi(u) + \phi(v) < \phi(0)$, since the derivatives do not exist on the set $\{(x,y) : \phi(x) + \phi(y) = \phi(0)\}$ where according to Genest and MacKay (1986a), the copula has a singular part. The probability of this set is $-\phi(0)/\phi'(0)$.

Indeed, if we integrate the density c over its domain using the transformation

$$z = C(u,v) \text{ and } w = u,$$

then, for $\mathcal{D} = \{(u,v) : \phi(u) + \phi(v) < \phi(0)\}$ the domain of c, and $\mathcal{D}'$ the transformed domain,

$$\begin{aligned}
\int\int_{\mathcal{D}} c(u,v)\,\mathrm{d}u\,\mathrm{d}v &= \int\int_{\mathcal{D}'} c(u(z,w),v(z,w))\left|\frac{\partial(u,v)}{\partial(z,w)}\right|\,\mathrm{d}z\,\mathrm{d}w \\
&= \int\int_{\mathcal{D}'} c(u(z,w),v(z,w))\left|\frac{\partial(z,w)}{\partial(u,v)}\right|^{-1}\,\mathrm{d}z\,\mathrm{d}w \\
&= \int\int_{0<z<w<1} -\frac{\phi''(z)}{[\phi'(z)]^2}\phi'(w)\,\mathrm{d}z\,\mathrm{d}w,
\end{aligned}$$

because the corresponding Jacobian matrix has determinant

$$\frac{\partial(z,w)}{\partial(u,v)} = -\frac{\phi(v)}{\phi'(z)}.$$

Evaluation of the last integral yields:

$$\int_0^1 \phi'(w)\int_0^w -\frac{\phi''(z)}{[\phi'(z)]^2}\,\mathrm{d}w\,\mathrm{d}z = 1 + \frac{\phi(0)}{\phi'(0)},$$

which is less that 1 if and only if

$$\frac{\phi(0)}{\phi'(0)} \neq 0.$$

Therefore, the copula has a singular part on the set $\{(u,v) : (u,v) \notin \mathcal{D}\}$, that is $\{(u,v) : \phi(u) + \phi(v) = \phi(0)\}$. A pair (U,V) belongs to this set with probability $-\phi(0)/\phi'(0)$.

Suppose now that

$$\phi(t) = 1 - t, \tag{2.18}$$

which is strictly decreasing, convex and satisfies the boundary condition $\phi(1) = 0$. Then (2.17) yields

$$C(u,v) = \max\{(u+v-1), 0\}.$$

However, ϕ does not have second derivative and hence, copula C (which is the same as the Fréchet's lower bound in this case) does not have a proper density. Furthermore,

$$\frac{\phi(0)}{\phi'(0)} = -1$$

therefore, C is completely singular. This implies that any pair (U,V) belongs to the set

$$\{(U,V) : U + V = 1\}$$

with probability one, or equivalently $V = 1 - U$ with probability one. This is a simple example demonstrating how easy singular components of Archimedean copulas can be detected. It is also an illustration of a bivariate distribution that has absolutely continuous margins but the joint distribution contains a singular part.

One may notice, that generators of the form

$$t^{-\theta} - 1 \quad \text{or} \quad \frac{t^{-\theta} - 1}{\theta},$$

correspond to the same copula and the latter includes (2.18) for $\theta = -1$. To summarize, $\phi(t) = (t^{-\theta} - 1)/\theta$ for $\theta \in [-1, \infty] \backslash \{0\}$ generates the Clayton copula

$$C(u,v) = \max\left\{\left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}, 0\right\}.$$

The generator is strict (i.e., $\phi(0) = \infty$) when $\theta > 0$. For $-1 \leq \theta < 0$, ϕ is non-strict and the distribution has a singular component only when $\theta = -1$. Otherwise $\phi'(0) = -\infty$ and the Clayton family is absolutely continuous.

Theorem 2.4.2. *Let C be an Archimedean copula with generator ϕ. Then, for all $0 < u, v, w < 1$,*

(a) C is symmetric, i.e., $C(u, v) = C(v, u)$

(b) C is associative, i.e., $C(C(u, v), w) = C(u, C(v, w))$

The term 'Archimedean' was first appear in Ling (1965) and is of algebraic origin. Specifically, if we consider C as an operator between two points u, v from $(0, 1)$ and denote $u_C^n = C(u, u^{n-1})$ with $u_C^1 = u$ for n a positive integer, then C satisfies the version of the Archimedean axiom that 'for all u,v, it exists a natural $n : u_C^n < v$. The proof can be found in Nelsen (1999).

Archimedean copulas are of interest because of the simplicity of their construction, as the univariate function ϕ determines C, but also because of their mathematical tractability. Furthermore, they can capture a variety of dependence structures and hence have become a valuable tool in data modelling.

It is straightforward to see that the independence copula $C^{\perp\!\!\!\perp}$, belongs to the Archimedean class with generator $\phi(t) = -\log(u)$. We should note that although the variety of models included in the Archimedean family is vast, it is only a small subset of the set of all copulas. The popularity of Archimedean copulas is mainly due to reasons of mathematical convenience. An example of an important non-Archimedean copula is the normal copula. In this thesis however we focus on the estimation of bivariate Archimedean copulas.

The following theorem, due to Genest and MacKay (1986b), describes a simple criterion to verify whether a particular copula belongs to the Archimedean class.

Theorem 2.4.3. *A copula C is Archimedean if and only if there exists a mapping $f : (0, 1) \to (0, \infty)$ such that:*

$$f(v)\frac{\partial}{\partial u}C(u, v) = f(u)\frac{\partial}{\partial v}C(u, v)$$

for all $0 < u, v < 1$. The generator ϕ is given (up to a constant) by:

$$\phi(t) = \int_t^1 f(z)\ \mathrm{d}z$$

Example 6. *The family of Farlie-Gumbel-Morgenstern (FGM) is defined by:*

$$C(u,v) = uv[1+\theta(1-u)(1-v)], \quad |\theta| \leq 1.$$

Then

$$\frac{\partial}{\partial u}C(u,v) = v[1+\theta(1-v)(1-2u)],$$

and

$$\frac{\frac{\partial}{\partial u}C(u,v)}{\frac{\partial}{\partial v}C(u,v)} = \frac{v[1+\theta(1-v)(1-2u)]}{u[1+\theta(1-u)(1-2v)]},$$

which is not in the form

$$\frac{f(u)}{f(v)},$$

except when $\theta = 0$ which gives $f(z) = 1/z$ and $C = C^{\perp\!\!\!\perp}$.

Limiting cases, where the Fréchet bounds C^-, C^+ or even the independence copula $C^{\perp\!\!\!\perp}$ (as in the last example) appear as members of a particular Archimedean family for certain parameter values, are easy to determine with the following theorems due to Genest and MacKay (1986b).

Theorem 2.4.4. *Let Φ be the set of all continuous strictly decreasing, convex functions*

$$\phi : [0,1] \to [0,\infty]$$

such as $\phi(1) = 0$ and C_θ be a family of Archimedean copulas with differentiable generators $\phi_\theta \in \Phi$. Then $C = \lim C_\theta$ is an Archimedean copula if and only if there exists a function $\phi \in \Phi$ such that for all s,t in $(0,1)$

$$\lim_\theta \frac{\phi_\theta(s)}{\phi'_\theta(t)} = \frac{\phi(s)}{\phi'(t)}$$

where the limit is taken as θ tends to an endpoint its domain.

Example 7. *Let ϕ_θ be the generator of the Clayton copula,*

$$\phi_\theta(t) = \frac{t^{-\theta}-1}{\theta}, \quad \theta \in [-1,\infty)\backslash\{0\},$$

Then,

$$\lim_{\theta\to-1}\frac{\phi_\theta(s)}{\phi'_\theta(t)} = -\lim_{\theta\to-1}\frac{t^{\theta+1}(s^{-\theta}-1)}{\theta} = s-1,$$

and since

$$\frac{\phi(s)}{\phi'(t)} = s - 1$$

for ϕ the generator of the the lower Fréchet bound C^-, $\phi(t) = 1 - t$, , we verify that $C_{-1} = C^-$.

For the same family, using L'Hôspital's rule, we also have:

$$\lim_{\theta \to 0} \frac{\phi_\theta(s)}{\phi'_\theta(t)} = \lim_{\theta \to 0} \frac{s^{-\theta} \log s}{t^{-\theta-1}(1 - \theta \log t)} = t \log s,$$

hence, $C_0 = C^{\perp\!\!\!\perp}$.

Theorem 2.4.5. *Let C_θ be a family of Archimedean copulas with differentiable generators $\phi_\theta \in \Phi$. Then, $\lim C_\theta(u, v) = \min(u, v) = C^+(u, v)$, if and only if*

$$\lim_{\theta} \frac{\phi_\theta(t)}{\phi'_\theta(t)} = 0,$$

for all t in $(0, 1)$ and as θ tends to an endpoint of its domain.

Example 8. *For the Clayton family and using L'Hôspital's rule again, it is not difficult to show that*

$$\lim_{\theta \to \infty} \frac{\phi_\theta(t)}{\phi'_\theta(t)} = \lim_{\theta \to \infty} \frac{t^{\theta+1} - t}{\theta} = \lim_{\theta \to \infty} t^{\theta+1} \log t = 0.$$

Thus, the upper Fréchet bound $C^+(u, v) = \min(u, v)$, is a special case of the Clayton family achieved as $\theta \to \infty$.

All these cases are typical examples demonstrating how uncomplicated the calculations involving members of the Archimedean class are. In addition, we saw in (2.8) that the evaluation of Kendall's τ requires a double integration operation which in general, may not have an analytic form. For Archimedean copulas however, the situation is simpler and Kendall's τ can be evaluated directly from the generator ϕ. We will also see that the tail dependence index can be expressed as a function of ϕ. The first theorem that follows is due to Genest and MacKay (1986b) and the second is from Joe (1997).

Theorem 2.4.6. *Let (U, V) be a pair of random variables whose joint distribution is some Archimedean copula $C(u, v) = \phi^{-1}(\phi(u) + \phi(v))$ for some $\phi \in \Phi$. Then,*

$$\tau(U, V) = 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} \, \mathrm{d}t + 1. \tag{2.19}$$

Example 9. *Consider again the Clayton family. Evaluation of the integral (2.19) yields:*

$$\begin{aligned}\tau &= 1 + 4\int_0^1 \frac{t^{\theta+1} - t}{\theta}\,\mathrm{d}t \\ &= 1 + \frac{4}{\theta}\left(\frac{1}{\theta+2} - \frac{1}{2}\right) \\ &= \frac{\theta}{\theta+2}, \quad \theta \in [-1,\infty)\backslash\{0\}. \end{aligned} \tag{2.20}$$

Theorem 2.4.7. *If $(\phi^{-1})'(0)$ is finite, the copula:*

$$C(u,v) = \phi^{-1}(\phi(u) + \phi(v))$$

does not have upper tail dependence. If C has upper tail dependence, then $(\phi^{-1})'(0) = -\infty$ and the tail dependence parameter is:

$$\lambda_U = 2 - 2\lim_{s\to 0}\frac{(\phi^{-1})'(2s)}{(\phi^{-1})'(s)}$$

Similarly, the lower tail dependence parameter is equal to:

$$\lambda_L = 2\lim_{s\to\infty}\frac{(\phi^{-1})'(2s)}{(\phi^{-1})'(s)}$$

Example 10. *For the Clayton family, we can verify the result in (2.13). Indeed,*

$$\begin{aligned}\lambda_L &= 2\lim_{s\to\infty}\frac{(\phi^{-1})'(2s)}{(\phi^{-1})'(s)} \\ &= 2\lim_{s\to\infty}\frac{(1+2\theta s)^{\frac{1}{\theta}-1}}{(1+\theta s)^{\frac{1}{\theta}-1}} \\ &= 2^{-\frac{1}{\theta}}. \end{aligned}$$

So far, for illustrations purposes we have mainly used the Clayton copula. This one-parameter family, as the name suggests, was introduced by Clayton (1978) in a context to investigate association in multivariate life tables and is discussed in Cook and Johnson (1981) and Oakes (1982). In literature, it is also encountered with the name Kimeldorf-Sampson family, after Kimeldorf and Sampson (1982). We summarize the main characteristics of this and some other widely used one-parameter Archimedean copulas in the following paragraph.

1. CLAYTON FAMILY. We have seen already, that the generator of this copula is
$$\phi(t) = \frac{t^{-\theta} - 1}{\theta}$$
and is associated with the distribution
$$C(u, v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}, \quad \theta > 0.$$
We can extend the copula and include negative parameter values by taking:
$$C(u, v) = \max\left\{\left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}, 0\right\}, \quad \theta \in [-1, \infty)\backslash\{0\}.$$
with the cost of the presence of a singular component in C. This distribution, when θ tends to -1, converges to $C^{-}(u, v) = \max\{u + v - 1, 0\}$ and as θ approaches to $+\infty$ the distribution becomes $C^{+}(u, v) = \min(u, v)$. The independence copula $C^{\perp\!\!\!\perp}$, is attained as $\theta \to 0$. Kendall's coefficient τ is
$$\tau = \frac{\theta}{\theta + 2}$$
and lies in the interval $[-1, 1]\backslash\{0\}$. Note that only the extended type is capable of capturing negative dependencies. This copula does not have upper tail dependence, but the lower tail dependence index exists and it is
$$\lambda_L = 2^{-\frac{1}{\theta}}.$$

2. FRANK FAMILY. It was introduced in a non-statistical context, by Frank (1979) who has shown that its members are the only distributions such that $C(1 - u, 1 - v) = 1 - u - v + C(u, v)$. Genest (1987) was the first to investigate the statistical properties of the family and to study the non-parametric estimation of the association parameter. The generator is
$$\phi(t) = -\log \frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \quad \theta \neq 0,$$
which gives the copula
$$C_\theta(u, v) = -\frac{1}{\theta} \log\left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right].$$

The dependence is maximal when $\theta \to +\infty$ and minimal as $\theta \to -\infty$ yielding, $C_{-\infty} = C^-$ and $C_{+\infty} = C^+$ respectively. The independence copula emerges as a special case when $\theta \to 0$ thus, $C_0 = C^{\perp\!\!\!\perp}$. The distribution has a peculiar property; it is symmetric around point $(\frac{1}{2}, \frac{1}{2})$ i.e., it is the same at $(1-u, 1-v)$ and (u, v). With regard to the tail dependence properties, the copula does not have neither lower nor upper tail dependence. Kendall's τ is equal to

$$\tau = 1 + 4\frac{D_1(\theta) - 1}{\theta}$$

which is mapped onto $[-1, 1]\backslash\{0\}$, where $D_1(\theta)$ is the Debye function of order one and remains undefined for the moment but it is presented in the next chapter (section (3.2.1)), where we discuss the estimation of the association parameter.

3. GUMBEL-HOUGAARD FAMILY. The generator of this copula is

$$\phi(t) = (-\log t)^\theta, \quad \theta \in [1, \infty).$$

and is named after its appearance in papers by Gumbel (1960a), Gumbel (1960b), Gumbel (1961) and Hougaard (1986a). The corresponding distribution is of the form:

$$C_\theta(u, v) = \exp\Big[-\big((-\log v)^\theta + (-\log v)^\theta\big)^{\frac{1}{\theta}}\Big].$$

Extreme members of the family are the upper Fréchet bound as θ approaches $+\infty$, thus $C_{+\infty} = C^+$, while the independence copula is achieved for $\theta = 1$, hence $C_1 = C^{\perp\!\!\!\perp}$. Kendall's index is equal to

$$\tau = \frac{\theta - 1}{\theta}$$

and varies between 0 and 1, and therefore, the copula can not describe negative association. Furthermore, the family possesses upper tail dependence,

$$\lambda_U = 2 - 2^{\frac{1}{\theta}}.$$

but no lower tail dependence.

The interested reader will find a detailed presentation of a very wide selection of copulas in Hutchinson and Lai (1990) or Joe (1997); the latter is more focused on dependence functions and also includes some vector-parameter families. In the next chapter we present a two-parameter family, discuss some estimation procedures and study the fit on simulated and real-world data. Central role to these investigations has the following theorem from Genest and Rivest (1993) which shows that any Archimedean copula is characterized by the cumulative distribution function of the random variable $Z = C(U, V)$.

Proposition 2.4.8. *Let U, V two uniform random variables on $(0, 1)$ with joint distribution some Archimedean copula $C(u, v) = \phi^{-1}(\phi(u)+\phi(v))$, and consider the transformations:*

$$W = \frac{\phi(U)}{\phi(U) + \phi(V)} \quad \text{and} \quad Z = C(U, V)$$

Then:

- *W is a uniform random variable on $(0, 1)$,*
- *the distribution function K of Z is:*

$$K(z) = z - \frac{\phi(z)}{\phi'(z)},$$

- *Z and W are independent random variables.*

Proof. Let $S = \phi(U)$, $T = \phi(V)$. Their joint distribution is:

$$\begin{aligned}\mathbb{P}\big(S \le s, T \le t\big) &= \mathbb{P}\big(U \ge \phi^{-1}(s), V \ge \phi^{-1}(t)\big) \\ &= 1 - \mathbb{P}\big(U \le \phi^{-1}(s)\big) - \mathbb{P}\big(V \le \phi^{-1}(t)\big) + \mathbb{P}\big(U \le \phi^{-1}(s), V \le \phi^{-1}(t)\big) \\ &= 1 - \phi^{-1}(s) - \phi^{-1}(t) + \phi^{-1}(s + t)\end{aligned}$$

and the conditional:

$$\begin{aligned}\mathbb{P}\big(S \le s | T = t\big) &= \frac{\partial \mathbb{P}\big(S \le s, T \le t\big)}{\partial t} \frac{1}{f_T(t)} \\ &= 1 - \frac{\phi'(\phi^{-1}(w))}{\phi'(\phi^{-1}(w + t))}\end{aligned}$$

because the density $f_T(t)$ of T equals to $1/\phi'(\phi^{-1}(t))$ since the derivative ϕ' exists a.e. due to the convexity of ϕ.

Therefore, the joint distribution of the random variables $W = \frac{\phi(U)}{\phi(U)+\phi(V)}$ and $Z = C(U,V)$

$$\begin{aligned}
\mathbb{P}\big(W \le w, Z \le z\big) &= \mathbb{P}\big(S \le w(S+T), \phi^{-1}(S+T) \le z\big) \\
&= \mathbb{P}\big(\phi(z) - T \le S \le \frac{wT}{1-w}\big) \\
&= -\int_{(1-w)\phi(z)}^{\phi(z)} \frac{\mathbb{P}\big(\phi(z) - t \le S \le \frac{wt}{1-w} \big| T = t\big)}{\phi'(\phi^{-1}(t))}\,\mathrm{d}t \qquad (2.21) \\
&\quad -\int_{\phi(z)}^{\phi(0)} \frac{\mathbb{P}\big(S \le \frac{wt}{1-w} \big| T = t\big)}{\phi'(\phi^{-1}(t))}\,\mathrm{d}t \qquad (2.22)
\end{aligned}$$

and integral (2.21) is evaluated to:

$$\begin{aligned}
(2.21) &= -\int_{(1-w)\phi(z)}^{\phi(z)} \frac{\mathbb{P}(S \le wt/(1-w)|T=t) - \mathbb{P}(S \le \phi(z) - t|T=t)}{\phi'(\phi^{-1}(t))}\,\mathrm{d}t \\
&= -\int_{(1-w)\phi(z)}^{\phi(z)} \Big(\frac{1}{\phi'(z)} - \frac{1}{\phi'(\phi^{-1}(t/(1-w)))}\Big)\,\mathrm{d}t \\
&= (1-w)\Big[\phi^{-1}\Big(\frac{\phi(z)}{1-w}\Big) - z\Big] - w\frac{\phi(z)}{\phi'(z)}
\end{aligned}$$

and

$$\begin{aligned}
(2.22) &= \int_{\phi(z)}^{\phi(0)} \mathbb{P}\Big(S \le \frac{wt}{1-w}\Big|T=t\Big)\frac{-1}{\phi'(\phi^{-1}(t))}\,\mathrm{d}t \\
&= -\int_{\phi(z)}^{\phi(0)} \Big(\frac{1}{\phi'(\phi^{-1}(t))} - \frac{1}{\phi'(\phi^{-1}(t/(1-w)))}\Big)\,\mathrm{d}t \\
&= z + (1-w)\Big[\phi^{-1}\Big(\frac{\phi(0)}{1-w}\Big) - \phi^{-1}\frac{\phi(z)}{1-w}\Big)\Big]\,\mathrm{d}t \\
&= z - (1-w)\phi^{-1}\Big(\frac{\phi(z)}{1-z}\Big)
\end{aligned}$$

and finally,

$$\mathbb{P}\big(W \le w, Z \le z\big) = w\Big(z - \frac{\phi(z)}{\phi'(z)}\Big)$$

for all points w and almost all z in $(0,1)$. It follows now that W is uniformly distributed on $(0,1)$ whereas

$$\mathbb{P}(Z \le z) = z - \frac{\phi(z)}{\phi'(z)}$$

and the variables Z, W are independent. □

Following the proposition and given knowledge of K, one could recover ϕ by solving the differential equation:

$$\frac{\phi(z)}{\phi'(z)} = z - K(z)$$

which yields:

$$\phi(z) = \exp\Big(\int_{z_0}^{z} \frac{1}{t - K(t)}\, \mathrm{d}t\Big) \tag{2.23}$$

for any point z_0 in $(0,1)$. When $K(t) > t$ for all t, the function in (2.23) is convex and decreasing. The boundary condition $\phi(1) = 0$ holds since:

$$0 \leq -\frac{1}{t-1} \leq -\frac{1}{t - K(t)}$$

and the integral of the middle term diverges on $(z_0, 1)$. Hence, the expression in (2.23) defines a generator. The following proposition stated by Genest and Rivest (1993) formally describes sufficient and necessary conditions for a function of the form (2.23) to be a generator.

Proposition 2.4.9. *Let U, V be uniform random variables on $(0,1)$ with joint distribution $C(u,v)$. Let $K(z) = \mathbb{P}\big(C(U,V) \leq z\big)$ and $K(z^-) = lim_{t \to z^-} K(t)$. The function ϕ defined in (2.23) is convex and decreasing and satisfies $\phi(1) = 0$ if and only if $K(z^-) > z$ for all points z in $(0,1)$.*

Note that copula C in the last proposition is not restricted to be an Archimedean one. It defines however a function of the form (2.23) which, when $K(z) > z$, is a generator and specifies an Archimedean copula C_ϕ. Hence, as Genest and Rivest (1993) mention, function (2.23) *can be used to define in some sense the 'projection' of (almost) any dependence function C within the class of Archimedean copulas.* A possible extension, which has not been investigated in literature and proposed here by the author may concern the construction of a metric

$$\sup_{u,v \in (0,1)} \big\{ |C(u,v) - C_\phi(u,v)| \big\}$$

to measure the distance of any copula C from the Archimedean class, given that $K(z) > z$ for all z in (0,1) and ϕ as in formula (2.23).

2.5 Construction of Copulas.

A general approach for the derivation of multivariate distributions has been proposed by Marshall and Olkin (1988). A special case of this approach concerns the derivation of Archimedean families.

Assume that G and H are univariate distributions and H has support on $(0, \infty)$. Then the function,

$$F(x) = \int G^{\theta}(x) \, \mathrm{d}H(\theta) \tag{2.24}$$

is a distribution function for $\theta > 0$. It is also interesting to see that for any distributions F, H there exists a distribution G such that equation (2.24) holds. Indeed, F is rewritten as:

$$\begin{aligned} F(x) &= \int G^{\theta}(x) \, \mathrm{d}H(\theta) \\ &= \int e^{\theta \log G(x)} \, \mathrm{d}H(\theta) \\ &= \psi(-\log G(x)) \end{aligned}$$

Hence, $G(x) = \exp\big(-\psi^{-1}(F(x))\big)$, where ψ is the Laplace transformation[2] of G.

Assume now that H is a bivariate distribution with margins H_1, H_2 and support on $(0, \infty) \times (0, \infty)$ and let:

$$F(x_1, x_2) = \int \int G_1^{\theta_1}(x_1) G_2^{\theta_2}(x_1) \, \mathrm{d}H(\theta_1, \theta_2)$$

where G_1, G_2 distribution functions and $\theta_1, \theta_2 > 0$. Then the margins of F are

$$F_1(x) = \int G_1^{\theta_1}(x) \, \mathrm{d}H_1(\theta_1) \qquad F_2(x) = \int G_2^{\theta_2}(x) \, \mathrm{d}H_2(\theta_2)$$

This is also true in the other direction:

Given distributions $G_i, i = 1, 2$ and Laplace transformations $\psi_i, i = 1, 2$ of some distributions $H_i, i = 1, 2$ we can define the following distributions:

$$F_i(x) = \psi_i(-\log G_i(x)), \quad i = 1, 2$$

[2]The Laplace transformation of a function F defined on $[0, \infty)$ is the function ψ defined by $\psi(\lambda) = \int_0^{\infty} e^{-\lambda t} \, \mathrm{d}F(t)$

It follows now that if H is a bivariate distribution on $(0, \infty)$ with margins H_1, H_2 then:

$$F(x_1, x_2) = \int \int G_1^{\theta_1}(x_1) G_2^{\theta_2}(x_2) \, \mathrm{d}H(\theta_1, \theta_2) \tag{2.25}$$

is a distribution function with margins F_1, F_2.

That proposes an alternative way to construct bivariate distributions, whereas the extension to higher dimensions is straightforward. Another observation concerns the integrand $G_1^{\theta_1}(x_1) G_2^{\theta_2}(x_2)$ in (2.25) which can be replaced by any distribution with margins $G_1^{\theta_1}, G_2^{\theta_2}$

We summarize with the following theorem by Marshall and Olkin (1988)

Theorem 2.5.1. *Let F_1, F_2 be univariate distributions and let H be a bivariate distribution such that $\bar{H}(0,0) = \mathbb{P}(X_1 > 0, X_2 > 0) = 1$ with univariate margins H_1, H_2. If C is a copula and $G_i(x) = \exp(-\psi_i^{-1}(F_i(x))), i = 1, 2$ where $\psi_i, i = 1, 2$ is the Laplace transformation of H_1, H_2 respectively then*

$$F(x_1, x_2) = \int \int C\big(G_1^{\theta_1}(x_1), G_2^{\theta_2}(x_2)\big) \, \mathrm{d}H(\theta_1, \theta_2)$$

is a bivariate distribution with margins F_1, F_2.

The following corollary demonstrates the derivation of Archimedean Copulas.

Corollary 2.5.2. *In theorem (2.5.1) let C be the product copula, i.e., $C(u, v) = u\,v$ and H be the upper Fréchet bound, $H(\theta_1, \theta_2) = \min\big(H_1(\theta_1), H_2(\theta_2)\big)$ with equal margins. Hence:*

$$\begin{aligned} F(x_1, x_2) &= \int G_1^{\theta}(x_1) G_2^{\theta}(x_2) \, \mathrm{d}H_1(\theta) \\ &= \psi\big[\psi^{-1}(F_1(x_1)) + \psi^{-1}(F_2(x_2))\big] \end{aligned}$$

where ψ is the Laplace transformation of the equal margins $G_i, i = 1, 2$.

Inverses of Laplace transformations can now serve as generators of Archimedean copulas. We remind that we call Archimedean every copula C, which has the following functional form:

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v))).$$

One can now see that the previous equation holds for a wider class of functions than inverses of Laplace transformations.

Briefly, the construction of bivariate distributions on $\mathbb{R}^2$ can be described by the following steps:

1. Choose univariate distributions F_1, F_2.
2. Choose a bivariate distribution H (of a positive random variable) with margins H_1, H_2.
3. Define $G_i(x_i) = \exp(-\psi_i^{-1}(F_i(x_i)))$ where ϕ_i the Laplace transformation of $H_i, i = 1, 2$.
4. Choose a (bivariate) distribution C with margins $G_i, i = 1, 2$. item Then:

$$F(x_1, x_2) = \int\int C(G_1^{\theta_1}(x_1), G_2^{\theta_2}(x_2))\, \mathrm{d}H(\theta_1, \theta_2)$$

is a bivariate distribution with margins F_1, F_2.

Some comments:

- In step (1), if we choose the distributions F_1, F_2 to be uniform on $(0, 1)$, then the resulting bivariate family in step (4) is a copula.
- When C is the product copula and H the upper Fréchet bound $H(\theta_1, \theta_2) = \min(H_1(\theta_1), H_2(\theta_2))$ with equal margins, the resulting family is an Archimedean copula indexed by the same vector of parameters that indexes the (equal) margins.
- When C is indexed by a number α, H is the upper Fréchet bound, $H(\theta_1, \theta_2) = \min(H_1(\theta_1), H_2(\theta_2))$ with equal margins, with parameter β the resulting family is a copula parameterized by a vector of two parameters (α, β).

Example 11. ***Clayton Family.*** *Let C, H be as in corollary (2.5.2) and the margins G are $\Gamma(\alpha, 1)$. Hence, the Laplace transformations of G is: $\psi(z) = (1+z)^{-\alpha}$ and $\psi^{-1}(z) = z^{-\frac{1}{\alpha}} - 1$. The generated distribution F is*

$$\begin{aligned} F(u, v) &= \psi(\psi^{-1}(u) + \psi^{-1}(v)) \\ &= \psi(u^{-\frac{1}{\alpha}} + v^{-\frac{1}{\alpha}} - 2) \\ &= (u^{-\frac{1}{\alpha}} + v^{-\frac{1}{\alpha}} - 1)^{-\alpha} \end{aligned}$$

i.e., the Clayton family.

Example 12. *The $\frac{1}{2}$-stable distribution $G(x) = 2(1-\Phi(\frac{1}{\sqrt{x}}))$ has Laplace transformation $\psi(z) = e^{-\sqrt{2z}}$. Hence, $\psi^{-1}(z) = \frac{1}{2}(\log z)^2$ and*

$$\begin{aligned} F(u,v) &= \psi(\tfrac{1}{2}(\log u)^2 + \tfrac{1}{2}(\log v)^2) \\ &= \exp\Big[-\big((\log u)^2 + (\log v)^2\big)^{1/2}\Big] \end{aligned}$$

which defines the Gumbel family with parameter 2.

Example 13. *Hougaard (1986b) has shown that*

$$\psi(s) = \exp\Big[\tfrac{\delta}{\alpha}(\theta+s)^\alpha - \theta^\alpha\Big], \quad 0 < \alpha < 1, \delta > 0, \theta > 0$$

is the Laplace transformation of some distribution $\mathbb{P}_{\alpha,\delta,\theta}$

The inverse of the above transformation is:

$$\phi(t) = \psi^{-1}(t) = \big(\theta^\alpha - \tfrac{\alpha}{\delta}\log t\big)^{\frac{1}{\alpha}} - \theta$$

which, following the last theorem, can be used for the construction of distribution functions.

Special Cases:

1. If $\theta = 0, \delta = 1$ and as $\alpha \to 1$ then $\phi(t) \to -\log t$ (which generates the product copula C(u,v) = uv and implies independence between the margins).

2. If $\theta = 0$ and $\delta = \alpha$ then $\phi(t) = (-\log t)^{1/\alpha}$, yielding the Gumbel-Hougaard family.

3. If $\alpha \to 0$, then the Laplace transformation $\psi(s) \to \big(\frac{\theta}{\theta+s}\big)^\delta$. Hence, ϕ is the Clayton family generator $\phi(t) = \theta(t^{-1/\delta} - 1)$.

4. **Log-copula.** If $\theta = 1$ then $\phi(t) = (1 - \frac{\alpha}{\delta}\log t)^{1/\alpha} - 1$, and if we set $\alpha^* + 1 = \frac{1}{\alpha}$ for some $\alpha^* \geq 0$ then

$$\phi(t) = \Big(1 - \frac{1}{\delta(\alpha^*+1)}\log t\Big)^{\alpha^*+1} - 1$$

which corresponds to the family of copulas mentioned by Genest and Rivest (1993) with the name *log-copula* and also includes the independence copula $C(u,v) = uv$ as a special case for $\alpha^* = 0$ and $\delta = 1$.

Example 14. *Assume now, that:*

- *C is the Gumbel-Hougaard copula:*

$$C(u,v) = \exp\big[-\big((-\log u)^{\beta} + (-\log u)^{\beta}\big)^{1/\beta}\big],$$

- *H is the upper Fréchet bound,*

$$H(\theta_1,\theta_2) = \min\big(H_1(\theta_1), H_2(\theta_2)\big)$$

with equal margins $H_1 = H_2 = H$ and

- *$H_1 \sim \Gamma(\alpha, 1)$, hence H_1 has Laplace transformation*

$$\psi(s) = (1+s)^{-1/\alpha}, \quad \alpha > 0$$

Corollary (2.5.2) yields:

$$\begin{aligned}
F(u,v) &= \int C\big(G_1^{\theta}(u), G_1^{\theta}(v)\big)\mathrm{d}H_1(\theta) \\
&= \int \exp\big[-\big((-\log G_1^{\theta}(u))^{\beta} + (-\log G_2^{\theta}(u)^{\beta}\big)^{1/\beta}\big]\ \mathrm{d}H_1(\theta_1) \\
&= \int \exp\big[-\theta\big((-\log G_1(u))^{\beta} + (-\log G_2(u)^{\beta}\big)^{1/\beta}\big]\ \mathrm{d}H_1(\theta_1) \\
&= \int \exp\big[-\theta\big((\psi^{-1}(u))^{\beta} + (\psi^{-1}(u)^{\beta}\big)^{1/\beta}\big]\ \mathrm{d}H_1(\theta_1) \\
&= \psi\big[\big((\psi^{-1}(u))^{\beta} + (\psi^{-1}(u)^{\beta}\big)^{1/\beta}\big] \\
&= \big[1 + \big((\psi^{-1}(u))^{\beta} + (\psi^{-1}(u)^{\beta}\big)^{1/\beta}\big]^{\frac{1}{\alpha}} \\
&= \big[1 + \big((u^{-\alpha}-1)^{\beta} + (v^{-\alpha}-1)^{\beta}\big)^{1/\beta}\big]^{\frac{1}{\alpha}}, \quad \alpha > 0,\ \beta > 1.
\end{aligned}$$

We will encounter this family of copulas in the remainder of this thesis where it is used to illustrate the estimation procedures presented herein and demonstrate that its performance is very encouraging for financial data modelling.

2.6 Other topics - Markov Processes.

Conditional dependence of two random variables can be characterized by the properties of their copula. Central to this investigation is a theorem stated by Darsow et al. (1992) based on an operator between copulas and introduces an alternative way for the construction of Markov processes.

Lemma 2.6.1. *If X and Y are random variables with copula C then:*

$$\mathbb{P}(X \leq x \,|\, Y = y) = \frac{\partial C\big(F(x), G(y)\big)}{\partial G(y)}$$

$$\mathbb{P}(Y \leq y \,|\, X = x) = \frac{\partial C\big(F(x), G(y)\big)}{\partial G(x)}$$

To avoid mathematical complexity we define the following operator:

Definition 5. *If $\mathcal{C}$ is the set of (two dimensional) copulas and $C_1, C_2 \in \mathcal{C}$ then define the product operator denoted by * as:*

$$(C_1 * C_2)(u, v) := \int_{(0,1)} \frac{\partial C_1(u, \omega)}{\partial \omega} \frac{\partial C_2(\omega, v)}{\partial \omega} \, \mathrm{d}\omega$$

for every $u, v \in (0, 1)$.

It follows now that:

Corollary 2.6.2. *If X, Y, Z are random variables and X, Z are independent conditionally on Y then:*

$$C_{XZ} = C_{XY} * C_{YZ}$$

Proof.

$$\begin{aligned} C_{XY} * C_{YZ}(x, z) &= \int \frac{\partial C\big(F_x(x), F_y(y)\big)}{\partial F_y(y)} \frac{\partial C\big(F_y(y), F_z(z)\big)}{\partial F_y(y)} \, \mathrm{d}F_y \\ &= \int \mathbb{P}(X \leq x \,|\, Y = y)\, \mathbb{P}(Z \leq z \,|\, Y = y) \, \mathrm{d}F_y \\ &= \int \mathbb{P}(X \leq x, Z \leq z \,|\, Y = y) \, \mathrm{d}F_y \\ &= \mathbb{P}(X \leq x, Z \leq z) = C_{XZ}(x, z) \end{aligned}$$

□

Theorem 2.6.3. *Let $\{X_t\}_{t>0}$ be a stochastic process and $C_{s,t}$ denote the copula of X_s and X_t. The following are equivalent:*

- *The transition probabilities $p_{s,t}(x, A) = \mathbb{P}(x_t \in A \,|\, X_s = s)$ satisfy the Chapman-Kolmogorov equations:*

$$p_{s,t}(x, A) = \int_{\mathbb{R}} p_{s,\theta}(x, \mathrm{d}y)\, p(y, A)$$

for all $s < \theta < t$ and $x \in \mathbb{R}$

- *For all* $s < \theta < t$,

$$C_{s,t} = C_{s,\theta} * C_{\theta,t} \tag{2.26}$$

What is demonstrated in the last theorem is an alternative way to construct Markov dependent stochastic processes. Instead of specifying the initial and the transition probabilities one can be based on a copula satisfying expression (2.26) and arbitrary chosen margins.

Example 15. *A brownian motion is a gaussian process* $\{X_t\}_{t>0}$ *with covariance matrix* $Cov(X_t, X_s) = \min(t, s)$ *and* $X_0 = 0$. *It follows that:*

$$X_t \mid X_s = x \sim N(x, t - s)$$

Thus:

$$\mathbb{P}(X_t \leq u \mid X_s = x) = \Phi\Big(\frac{y - x}{\sqrt{t - s}}\Big).$$

However,

$$p_{s,t}(x, (-\infty, y)] = \mathbb{P}(X_t \leq y \mid X_s = x) = \frac{\partial C\big(F_t(y), F_s(x)\big)}{\partial F_s(x)}$$

If we denote $F_s(x) = u$, $F_t(y) = v$,

$$\frac{\partial}{\partial u} C(u, v) = \Phi\Big(\frac{F_t^{-1}(v) - F_s^{-1}(u)}{\sqrt{t - s}}\Big)$$

Hence:

$$C(u, v) = \int_0^u \Phi\Big(\frac{F_t^{-1}(v) - F_s^{-1}(\omega)}{\sqrt{t - s}}\Big)\, \mathrm{d}\omega$$

Because the process is a gaussian one:

$$F_t(y) = \Phi\Big(\frac{y}{\sqrt{t}}\Big) \quad \text{and} \quad F_s(x) = \Phi\Big(\frac{x}{\sqrt{s}}\Big)$$

and the last integral becomes:

$$C(u, v) = \int_0^u \Phi\Big(\frac{\sqrt{t}\Phi^{-1}(u) - \sqrt{s}\Phi^{-1}(\omega)}{\sqrt{t - s}}\Big)\, \mathrm{d}\omega$$

This function describes a family of copulas, indexed by t, s and by construction satisfies (2.26) and theorem (2.6.3) implies that it can be used to generate Markov Processes when arbitrary margins are fitted. An example of a Markov process follows: Let U, V random variables whose joint distribution is as above. Then:

$$\mathbb{P}(U \leq u \,|\, V = v) = \frac{\partial C(u,v)}{\partial u} = \Phi\Big(\frac{\sqrt{t}\Phi^{-1}(u) - \sqrt{s}\Phi^{-1}(v)}{\sqrt{t-s}}\Big)$$

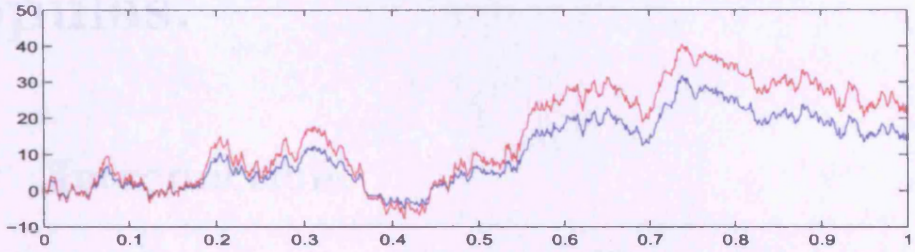
Assume that $U = F_t(X_t)$ and $V = F_s(X_s)$ for some random variables X_t, X_s with margins $F_t, F_s, t > s \geq 0$ and derive:

$$X_t = F_t^{-1}\Big[\Phi\Big(\sqrt{\frac{s}{t}}\,\Phi^{-1}(F_s(X_s)) + \sqrt{\frac{t-s}{t}}\,\varepsilon\Big)\Big],$$

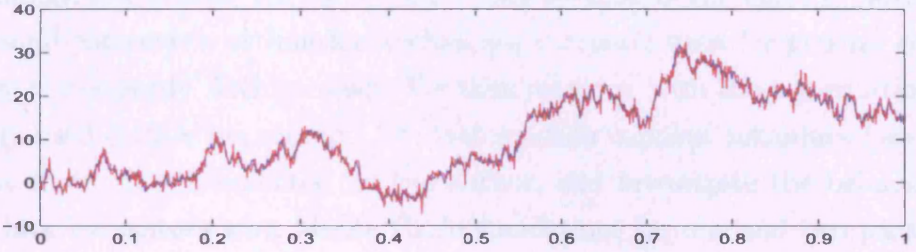
where ε random from $N(0,1)$. Different margins in the last expression will generate a Markov process whose dependence structure is the same as the one contained in a Brownian motion. Apparently normal margins, $X_t \sim N(0,t)$ yield the Brownian motion. Indeed, conditionally on $X_s = x_s$ gives:

$$\begin{aligned} X_t &= \sqrt{t}\,\Phi^{-1}\Big[\Phi\Big(\sqrt{\frac{s}{t}}\,\Phi^{-1}(\Phi(\frac{x_s}{\sqrt{s}})) + \sqrt{\frac{t-s}{t}}\,\varepsilon\Big)\Big] \\ &= x_s + \sqrt{t-s}\,\varepsilon \end{aligned}$$

Assume now the margins $X_t = \sqrt{\frac{(\nu-2)t}{\nu}}$ where $Y \sim t_\nu, \nu > 2$. A realization of the derived process for different values of ν is illustrated in figure (2.6). As expected, the process rapidly converges to a Brownian motion as ν increases.



(a) Margins follow Student's distribution with 3 degrees of freedom.



(b) Margins follow Student's distribution with 10 degrees of freedom.

Figure 2.6: Markov process with a Brownian motion dependence structure but different margins (in red colour). The process converges to a Brownian motion (blue colour) as the margins tend to normality.

Chapter 3

Estimation of Archimedean copulas.

3.1 Introduction.

This chapter aims to present some new parametric estimation procedures for Archimedean copulas. We start with a brief revision of the current parametric and semi-parametric estimation techniques currently used for general copulas i.e., not necessarily Archimedean. We then continue with the presentation of a widely used estimation method for Archimedean copulas, introduce two alternative methods implemented by the author, and investigate the behaviour of the three estimators with Monte Carlo simulations for one and two parameter families of copulas. An illustrative example with financial data concludes the chapter.

In the remainder we assume that a random sample $(X_1, Y_1), (X_2, Y_2), \ldots,$ (X_n, Y_n) has been drawn from a distribution H with margins F, G and copula C. We denote the respective densities with the appropriate lowercase letters. Let also $\theta_1, \theta_2, \alpha$ be the parameters of the margins F, G and the copula C respectively. We use the shorthand notation $(\boldsymbol{x}, \boldsymbol{y})$ for the observed sample $(x_1, y_1), (x_2, y_2) \ldots, (x_n, y_n)$.

Usually, estimation problems concerning a copula C arise when a joint distribution H is expressed as a function of the margins F, G

$$H(\boldsymbol{x}, \boldsymbol{y} \mid \theta_1, \theta_2, \alpha) = C\big(F(\boldsymbol{x} \mid \theta_1), G(\boldsymbol{y} \mid \theta_2) \mid \alpha\big), \tag{3.1}$$

and the parametric form of the margins is unknown. In such cases, it is clear that traditional parametric methods such as maximization of the full likelihood

$$h(\theta_1, \theta_2, \alpha \mid \boldsymbol{x}, \boldsymbol{y}) = c(F(\theta_1 \mid \boldsymbol{x}), G(\theta_2 \mid \boldsymbol{y}) \mid \alpha)\, f(\theta_1 \mid \boldsymbol{x})\, g(\theta_2 \mid \boldsymbol{y})$$

can not be applied. Likelihood methods involve simultaneous maximization of the model parameters in the joint distribution H. Under the assumption of known parametric families for the margins, representation (3.1) allows for procedures where the estimation can be carried out separately for F, G and H. in this way we minimize the computational burden for high dimensional problems where the full likelihood approach may become very demanding. Within this framework lies the method of Inference of Functions for Margins (IFM) proposed by Joe (1997). Briefly this approach can be summarized by

- Estimates $\theta_1^\dagger, \theta_2^\dagger$ for parameters θ_1, θ_2 of the margins F, G are obtained by separately maximizing the corresponding likelihoods $f(\theta_1 \mid \boldsymbol{x})$ and $g(\theta_2 \mid \boldsymbol{y})$ of the univariate margins.

- Conditionally on $\theta_1^\dagger, \theta_2^\dagger$, the pseudolikelihood

  $$L(\alpha \mid \theta_1^\dagger, \theta_2^\dagger) = h(\alpha \mid \boldsymbol{x}, \boldsymbol{y}, \theta_1^\dagger, \theta_2^\dagger)$$

  is maximised over α to obtain $\alpha^\dagger$.

The properties of the estimator $(\alpha_1^\dagger, \theta_1^\dagger, \theta_2^\dagger)$, such as consistency, efficiency and its asymptotic variance-covariance matrix and efficiency are detailed in Joe (1997).

Another method which can be considered as the semiparametric equivalent of the IFM method is presented in Genest et al. (1995). Empirical margins are used, which are scaled with a factor of $\frac{n}{n+1}$ to avoid, as the authors mention, '*difficulties arising from the potential unboundedness of the* $\log(c(x, y \mid \alpha))$ *as some of the* x, y *tend to one*'. The estimator for the copula parameter α is the value that maximizes the function:

$$l(\alpha|x) = \sum_{i=1}^{n} \log \left[c(\alpha \mid F^*(x_i), G^*(y_i)) \right] \tag{3.2}$$

where F^*, G^* are the scaled margins.

3.2 Estimation for Archimedean copulas.

For Archimedean copulas

$$C_\phi(u,v) = \phi^{-1}(\phi(u) + \phi(v)) \quad 0 < u, v < 1$$

where the generator ϕ is a convex, decreasing function such that $\phi(1) = 0$, Genest and Rivest (1993) have developed a fully nonparametric approach for the determination of ϕ and therefore of the copula C_ϕ. The following proposition, encountered in chapter 2 and restated here, is central to the approach:

Proposition 3.2.1. *Let U, V two uniform random variables on $(0,1)$ with joint distribution some Archimedean copula $C(u,v) = \phi^{-1}(\phi(u) + \phi(v))$. Consider the transformations:*

$$W = \frac{\phi(U)}{\phi(U) + \phi(V)} \quad \text{and} \quad Z = C(U,V).$$

Then:

- *W is a uniform random variable on $(0,1)$,*
- *the distribution function K of Z is:*

$$K(z) = z - \frac{\phi(z)}{\phi'(z)},$$

- *Z and W are independent random variables.*

In practice and when a sample of pairs $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ from some distribution H has been observed, one can estimate K, in a nonparametric manner, by the following two step procedure:

- construct the pseudobservations:

$$Z_i = \frac{\#\{(X_j, Y_j) : X_j < X_i, Y_j < Y_i\}}{n-1} \tag{3.3}$$

  where # denotes the cardinality of a set. This is possible because:

$$Z_i = C(U_i, V_i) = C(F(X_i), G(Y_i)) = \mathbb{P}(X \leq X_i, Y \leq Y_i)$$

  and hence, a realization of $Z_i = C(U_i, V_i)$ can be considered as the number of points componentwise less than or equal to (X_i, Y_i). The presence of $n-1$, instead of n in the denominator as normally expected, is justified later.

- For any point z in $(0,1)$, estimate $K = K(u)$ by:

$$K_n(z) = \frac{\sum_i I_{\{z \leq Z_i\}}}{n}$$

where $I_{\{x \in A\}}$ is the usual indicator function taking the value of one if $x \in A$ and zero otherwise.

Following (2.23), an estimator ϕ_n of the generator ϕ is given by

$$\phi_n(z) = \exp\Big(\int_{z_0}^{z} \frac{1}{t - K_n(t)} \mathrm{d}t\Big) \tag{3.4}$$

when $K_n(t) > t$ for all t in (0,1).

It is easy to observe that the pseudobservations $Z_i, i = 1 \ldots n$ are strongly related to the sample value of Kendall's tau, which is defined as the proportion of concordant pairs minus the proportion of disconcordant pairs in the sample. Specifically, if we denote:

$$I_{ij} = \begin{cases} 1 & \text{if } X_j < X_i \text{ and } Y_j < Y_i, \\ 0 & \text{otherwise,} \end{cases}$$

observations i and j are concordant if and only if $I_{ij} + I_{ji} = 1$. Hence, the number of concordant pairs is $\sum_{ij} I_{ij}$ and from the definition of Z_i:

$$\sum_{ij} I_{ij} = (n-1) \sum_i Z_i,$$

Therefore, τ, the sample value of Kendall's tau, is:

$$\tau = \frac{(n-1)\sum_i Z_i}{\frac{n(n-1)}{2}} - \Big(1 - \frac{(n-1)\sum_i Z_i}{\frac{n(n-1)}{2}}\Big) = 4\frac{\sum_i Z_i}{n} - 1$$

and if we denote $\bar{Z} = \frac{\sum_i Z_i}{n}$, it simplifies to:

$$\tau = 4\bar{Z} - 1 \tag{3.5}$$

which is the sample equivalent of the population identity

$$\tau = 4\mathbb{E}(Z) - 1 \tag{3.6}$$

directly derived from:

$$\tau = 4 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} H(x,y)\, \mathrm{d}H(x,y) - 1,$$

(Schweizer and Wolff, 1981). The use of $n-1$ as denominator in the definition of the pseudobservations in (3.3) aims to make the relation between τ and $\bar{Z}$ mathematically more elegant and to match the theoretical expression (3.6). Because of the relation between the pseudobservations and Kendall's τ, Nelsen et al. (2003) refer to the distribution K of the random variable $Z = C(U,V) = H(X,Y)$ as the *Kendall distribution function* of (X,Y). In the same work the authors also present a nice treatment of the properties of this class of distributions.

Whereas equation (3.4) provides a framework to construct a non-parameric estimator ϕ_n of the generator ϕ, in practice it is more convenient to use K_n for model selection, i.e. amongst different classes of generators choose the one whose member gives the best fit. Specifically, Genest and Rivest (1993) mention that *'it will be theoretically more meaningful - as well as computationally more convenient- to use K_n as a tool to help identify the parametric family of Archimedean copulas that provides the best possible fit to the data'*. A direct implication is that the nonparametric technique turns into a semiparametric one. One may consider this approach comparable to the one briefly described at the beginning of this chapter and involves empirical margins and then maximization of the pseudolikehood (3.2), because K_n is constructed in terms of an empirical estimator of the joint $H(x,y)$. The main difference however, is that the estimator is independent of the margins. Knowledge of the true marginals or even use of misspecifed ones leaves the estimator invariant; a property not shared amongst the other approaches that do not involve K. Then, conditionally that the parametric family of the generator is known, we choose the appropriate member ϕ_α based on the properties of $K_n(z)$.

Attention is now on the estimation of the association parameter α that indexes the generator ϕ and hence the Archimedean copula C_ϕ itself. Because of the relationship between K_n and τ, a natural way to proceed is to use statistics (3.5) and (3.6). Given a parametric family for ϕ, an estimator for α is constructed in such a manner that those statistics are equal. Equivalently and maybe more transparently, we can bypass the construction of the pseudosample and for a given measure of association, find the parameter value α for which the sample value of the statistic is the same as the theoretical one. This approach follows the spirit of the method of moments and has been investigated in Genest (1987) for the Frank copula which is indexed by a scalar. For families

with a vector parameter a natural extension is to find as many moments of the pseudosample as needed and then solve the corresponding system of equations.

We now present an alternative estimation method which is also based on the pseudobservations and their probabilistic properties. Our motivation stems from the fact that the method of moments does not seem to perform well when the copula parameter is a vector. This is documented in Genest and Rivest (1993) who state that there are '*doubts on the efficiency of extending the method-of-moments estimation procedure to handle multiparameter families, such as the log-copulas and suggests that alternative approaches should be investigated*'. We propose two competing methods:

- Minimization of the distance between the theoretical K and the empirical K_n. Thus the estimator for the copula parameter α is the value α^* that minimizes the L_2 distance:

$$\left(\sum_{i=1}^{n} \left(K(\alpha, z_i) - K_n(z_i)\right)^2\right)^{1/2} \tag{3.7}$$

- Maximum likelihood on the pseudobservations Z_i. The estimator is

$$\hat{\alpha} = \arg\max_{\alpha} \sum_{i=1}^{n} \log\left(k(\alpha|z_i)\right) \tag{3.8}$$

  where k is the density of K.

A diagrammatic representation of the proposed methods along with other traditional copula estimation methods is illustrated in figure 3.1.

Under the method-of-moments approach, the estimation can be handled using Kendall's tau or Spearman's rho. The first one has already been defined and a relationship with the pseudobservations has been established in (3.6); An identity useful for the expansion of the method for the multivariate case. The sample value of Spearman's rho is given by:

$$\varrho = 1 - 6\sum_{i=1}^{n} \frac{D_i^2}{n(n^2-1)}$$

where D_i is the difference between the rank of X_i and that of Y_i and n is the sample size, whereas its population counterpart is given by:

$$\rho = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left(H(x,y) - F(x)G(y)\right) \, \mathrm{d}F(x)\mathrm{d}G(y)$$

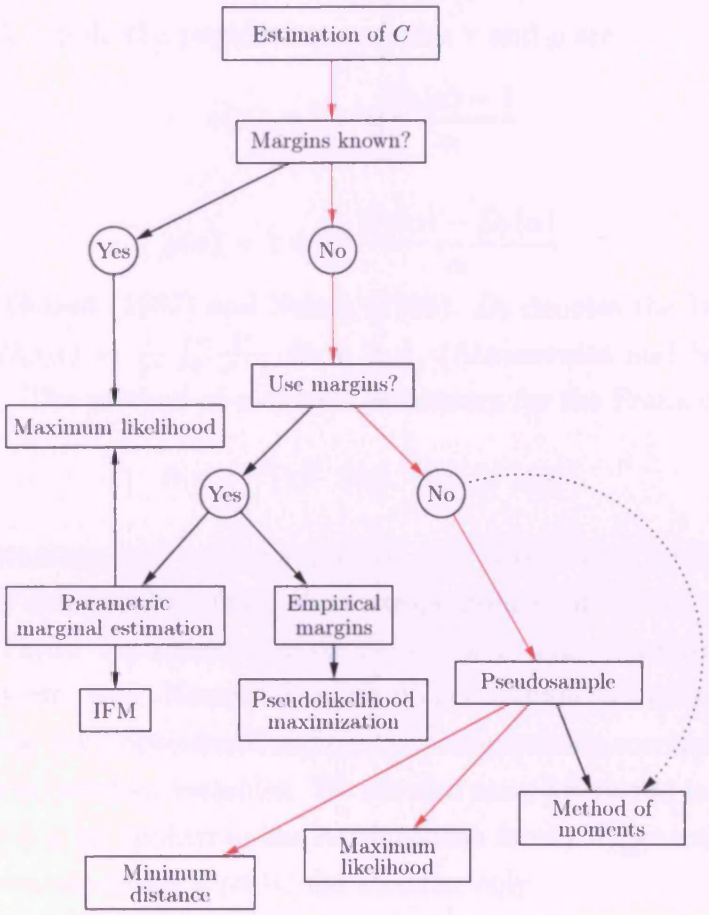


Figure 3.1: Estimation methods for Archimedean copulas. Red lines represent the approaches introduced here whereas black coloured ones are for the traditional methods. The dotted curve stands for a shortcut. The diagram makes clear the differences of the estimators proposed in this chapter from their competitors; most importantly the invariance of any marginal inference.

A very good treatment on issues concerning measures of dependence is given in Schweizer and Wolff (1981).

3.2.1 Application for a family indexed by a real parameter.

For the Frank copula the population statistics τ and ρ are:

$$\tau(\alpha) = 1 + 4\frac{D_1(a) - 1}{\alpha}$$

and

$$\rho(\alpha) = 1 + 12\frac{D_2(\alpha) - D_1(\alpha)}{\alpha}$$

respectively, Genest (1987) and Nelsen (1986). D_k denotes the Debye function of order k, $D_n(\alpha) = \frac{n}{\alpha^n}\int_0^\alpha \frac{t^n}{e^t-1}\,\mathrm{d}t, n \geq 1$, (Abramowitz and Stegun (1972), section 27.1). The method-of-moments estimators for the Frank copula are:

$$\tilde{\alpha} = \tau^{-1}(\tau) \quad \text{and} \quad \tilde{\alpha} = \rho^{-1}(\varrho)$$

These estimators can not be expressed in a closed form, and estimation is carried out by numerically solving the corresponding expressions. Genest (1987) proposed a Taylor expansion approximation of ρ, and demonstrated that it generally performs well. Nevertheless, we do not include this estimator into our studies because the approximation fails in cases of strong correlation ($\varrho > 0.7$) between the two random variables. We are also more interested in investigating methods that can be applied to the Archimedean family in general, rather than focusing on estimators for a particular member only.

For the Frank copula which has a generator of the form

$$\phi(z) = -\log\frac{e^{-\alpha z} - 1}{e^{-\alpha} - 1}, \quad \alpha \neq 0 \tag{3.9}$$

the distribution K of the random variable $Z = C(U, V)$ turns out to be:

$$K(z) = z + \frac{1 - e^{-\alpha z}}{\alpha}\log\frac{e^{-\alpha z} - 1}{e^{-\alpha} - 1}$$

with corresponding density

$$k(z) = -e^{\alpha z}\log\frac{e^{-\alpha z} - 1}{e^{-\alpha} - 1}$$

for any real α different than zero. When α tends to zero the limiting form of the density is

$$\lim_{\alpha \to 0} k(z) = -\log(z)$$

which, as expected, is also the expression for k that corresponds to independent margins U and V (with associated generator $\phi(z) = \log(z)$ itself, of the independence copula $C^{\perp}(u, v) = uv$). Graphs of the distribution and its density are included in figure 3.2.

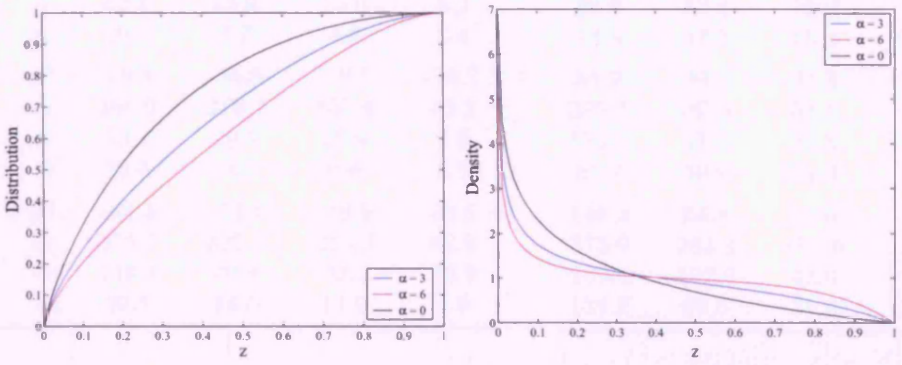


Figure 3.2: The distribution $K(z) = z - \phi(z)/\phi'(z)$ (left) and its density (right) where $\phi(z)$ is the generator (3.9) of the Frank copula for different values of α. The case where α tends to 0 corresponds to independent margins.

The estimators (3.8) and (3.7) for the Frank parametric family are:

$$\alpha^* = \arg\min_{\alpha} \left\{ \sum_{i=1}^{n} \left(z_i + \frac{1 - e^{-\alpha z_i}}{\alpha} \log \frac{e^{-\alpha z_i} - 1}{e^{-\alpha} - 1} - K_n(z_i) \right)^2 \right\}^{1/2} \tag{3.10}$$

and

$$\hat{\alpha} = \arg\max_{\alpha} \sum_{i=1}^{n} \log \left(-e^{\alpha z} \log \frac{e^{-\alpha z} - 1}{e^{-\alpha} - 1} \right) \tag{3.11}$$

In both cases, the optimal values can not be expressed in a closed form and numerical methods are used to obtain the solution. The investigation of the small sample behaviour of the estimators is accomplished in terms of a Monte Carlo simulation. A study of the bias and mean squared error of the competing estimators was carried out for three values of α providing a wide spectrum of correlation, varying from $\rho = 0.28$ to 0.71 (for $\alpha = 1.5$ to 6 respectively) and

for sample sizes $n = 10, 15, 25, 50$. For each value of α, 1000 samples of size n were generated and the estimated bias and mean squared error were recorded on table 3.1.

		Bias (× 100)				Mean squared error (× 10)			
α		$n = 10$	$n = 15$	$n = 25$	$n = 50$	$n = 10$	$n = 15$	$n = 25$	$n = 50$
1.5	α^*	-7.5	-20.3	-12.3	-6.6	68.7	35.4	18.8	9.5
	$\hat{\alpha}$	299.0	154.4	80.2	35.4	265.9	89.3	29.9	10.5
	$\bar{\alpha}$	42.8	15.0	10.0	5.2	96.6	37.9	19.9	9.0
	$\tilde{\alpha}$	29.8	7.7	4.8	3.4	74.8	33.3	18.4	8.7
3	α^*	-19.4	-35.8	-19.5	-10.2	85.0	39.3	21.3	11.0
	$\hat{\alpha}$	368.9	210.7	105.4	43.2	379.4	165.1	51.1	14.3
	$\bar{\alpha}$	73.7	29.5	19.9	9.9	132.9	47.6	24.3	11.0
	$\tilde{\alpha}$	38.6	13.7	9.6	5.5	87.7	39.9	21.9	10.5
6	α^*	-62.3	-73.4	-49.5	28.6	148.0	64.8	33.6	17.2
	$\hat{\alpha}$	375.7	322.5	215.3	82.9	373.6	283.2	157.6	39.7
	$\bar{\alpha}$	146.3	70.8	39.3	16.9	289.4	102.9	43.0	18.5
	$\tilde{\alpha}$	39.8	18.0	11.9	4.0	134.2	69.5	35.8	17.5

Table 3.1: Estimated Bias and Mean Squared Error from a Monte Carlo study for the estimators of α. The table shows results for our minimum distance and maximum likelihood estimators, $\alpha^*, \hat{\alpha}$ and the method of moments $\bar{\alpha}$ (based on Kendall's tau) and $\tilde{\alpha}$ (based on Spearman's rho).

The results for the method-of-moments estimators $\bar{\alpha}$ and $\tilde{\alpha}$ show great similarity with those from the original paper from Genest (1987).

The most interesting results of the simulation are as follows:

- Maximum estimation based on the pseudobservations does not seem to perform well as the corresponding estimated mean squared error is the largest across all sample sizes and all parameter values for α. This mainly results from the estimation error in the the construction of the pseudosample.

- Estimation based on minimizing the sum of the square differences between the theoretical and the empirical values underestimate α. The remaining three estimators show positive bias.

- Generally speaking the sum of squared differences estimator α^* appears to have the smallest mean squared error for almost all cases. This is surprising inasmuch as this estimator is based on the pseudosample, a feature shared by $\hat{\alpha}$ which does not exhibit similar performance.

3.2.2 Application for the estimation of a vector parameter.

In the remainder of this chapter we apply our proposed method to estimate a multiparameter Archimedean copula. This is where our contribution is mostly concentrated, since the method proposed by Genest and Rivest (1993) does not seem to cover this case in a satisfactory manner. In the case of a scalar parameter, our method does not lack in performance in terms of the mean squared error and when it competes against the method-of-moments. On the contrary, it appears from table 3.1 that in most cases it is at least as good as its competitors.

A straightforward way to construct multiparameter families is by composite generators as the following proposition shows:

Proposition 3.2.2. *Let ϕ be a generator. Then the functions:*

$$\phi_\alpha = \phi(t^\alpha) \quad \textit{and} \quad \phi_\beta = \big(\phi(t)\big)^\beta$$

and their composition:

$$(\phi_\alpha \circ \phi_\beta)(t) = \phi_\beta(\phi_\alpha(t)) = \big(\phi(t^\alpha)\big)^\beta$$

ia a class of generators for α in $(0, 1]$ and $\beta \geq 1$.

The domain of α can be extended to include all positive real numbers by imposing that $t\phi'(t)$ is nondecreasing on $(0, 1)$.

For example, from the very simple generator $\phi(t) = t^{-1} - 1$ we obtain $\phi_\alpha(t) = t^{-\alpha} - 1$ and finally the multiparameter generator:

$$\phi_{\alpha,\beta}(t) = (t^{-\alpha} - 1)^\beta \tag{3.12}$$

which is associated with the copula:

$$C(u, v) = \left[1 + \left((u^{-\alpha} - 1)^\beta + (v^{-\alpha} - 1)^\beta\right)^{1/\beta}\right]^{-1/\alpha} \quad \alpha > 0,\ \beta \geq 1 \tag{3.13}$$

Two parameter families provide us with a degree of flexibility since we are able of capturing more than one type of dependence, for example lower and upper tail dependence. The copula introduced in (3.13) has both upper and tail dependence,

$$\lambda_U = \lim_{u \to 1} \frac{\mathbb{P}(U > u, V > u)}{1 - u} = 2 - 2^{1/\beta} \tag{3.14}$$

and

$$\lambda_L = \lim_{u \to 0} \frac{\mathbb{P}(U \leq u, V \leq u)}{u} = 2^{-1/(\alpha\beta)} \tag{3.15}$$

It can also be seen, that copulas of the form (3.13) include other subfamilies of distributions. The Clayton family is obtained for $\alpha = 1$ whereas as β tends to infinity it reduces to the Gumbel family. The independence copula $C(u, v) = uv$ corresponds to $\beta \to \infty$ in which case $U = 1 - V$. When $\alpha > 0$ and $\beta \geq 1$ the density of the copula in 3.13 is given by

$$c(u,v) = \frac{t_1^{\beta}\, t_2^{\beta}\, (t_1^{\beta} + t_2^{\beta})^{-2+\frac{1}{\beta}} \left(1 + (t_1^{\beta} + t_2^{\beta})^{\frac{1}{\beta}}\right)^{-2-\frac{1}{\alpha}} \left(-\alpha + \alpha\beta + (t_1^{\beta} + t_2^{\beta})^{\frac{1}{\beta}}(1 + \alpha\beta)\right)}{u\, v\, (u^{\alpha} - 1)(v^{\alpha} - 1)}$$

for $t_1 = u^{-\alpha} - 1$ and $t_2 = v^{-\alpha} - 1$. When $\beta = 1$ the density is reduced to:

$$c(u,v) = \frac{(1+\alpha)\left(u^{-\alpha} + v^{-\alpha} - 1\right)^{-2-1/\alpha}}{(uv)^{\alpha+1}}.$$

Hence:

$$\lim_{\substack{\alpha \to 0 \\ \beta = 1}} c(u,v) = \frac{1}{uv} \lim_{\alpha \to 0} (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha} \tag{3.16}$$

and the limit in the RHS is rewritten as:

$$\lim_{\alpha \to 0} \exp\left(-\frac{1}{\alpha} \log(u^{-\alpha} + v^{-\alpha} - 1)\right).$$

The exponent tends to:

$$\lim_{\alpha \to 0} \frac{u^{-\alpha} \log u + v^{-\alpha} \log v}{u^{-\alpha} + v^{-\alpha} - 1} = \log(uv)$$

and finally (3.16) yields:

$$\lim_{\substack{\alpha \to 0 \\ \beta = 1}} c(u,v) = 1 \quad \text{for all } 0<u,v<1,$$

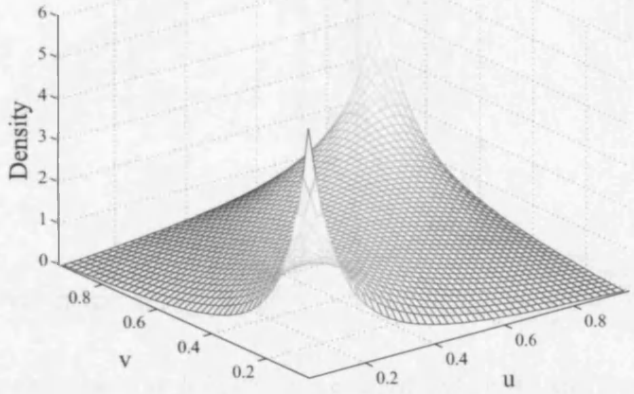


Figure 3.3: Density plot of the BB1 copula with $\alpha = 0.5, \beta = 1.5$.

as expected since this particular choice of parameter values corresponds to independent margins. The copula of the form (3.13) will be encountered on several occasions in this thesis. We name it after the labelling in Joe (1997) and call it the **BB1 family**. The graph of its density is given in figure 3.3.

The generator in (3.12) gives the following distribution K

$$K(z) = z - \frac{\phi(z)}{\phi'(z)} = z + \frac{z^{\alpha+1}(z^{-\alpha} - 1)}{\alpha\beta}, \tag{3.17}$$

with density

$$k(z|\alpha, \beta) = 1 + \frac{1}{\alpha\beta} - \frac{1+\alpha}{\alpha\beta} z^{\alpha}.$$

Plots of the distribution (3.17) and its density are included in figure 3.4. Following the identity $\tau = 4\mathbb{E}(Z) - 1$, simple algebraic manipulations give that

$$\tau = 4 \int_0^1 z \, \mathrm{d}K(z) - 1 = 4 \int_0^1 \frac{\phi(z)}{\phi'(z)} \, \mathrm{d}z + 1$$

which, for the generator in (3.12) implies that the population Kendall's tau is given by

$$\tau = 1 - \frac{2}{\beta(\alpha + 2)} \qquad \alpha > 0, \beta \geq 1,$$

It is straightforward to see that it is increasing in both α and β and reaches 1 when at least of the parameters approaches infinity, in which case the margins U, V are such that $U = V$. A plot is included in figure 3.5.

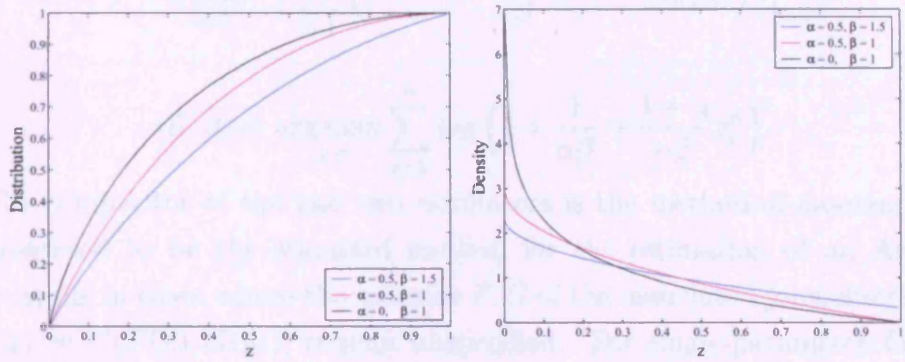


Figure 3.4: The distribution $K(z) = z - \phi(z)/\phi'(z)$ (left) and its density (right) for $\phi(z) = (z^{-\alpha} - 1)^{\beta}$ and various values for α, β. The red curve ($\alpha = 0.5, \beta = 1$) corresponds to the Clayton copula and the black coloured one ($\alpha \to 0, \beta = 1$) to independence.

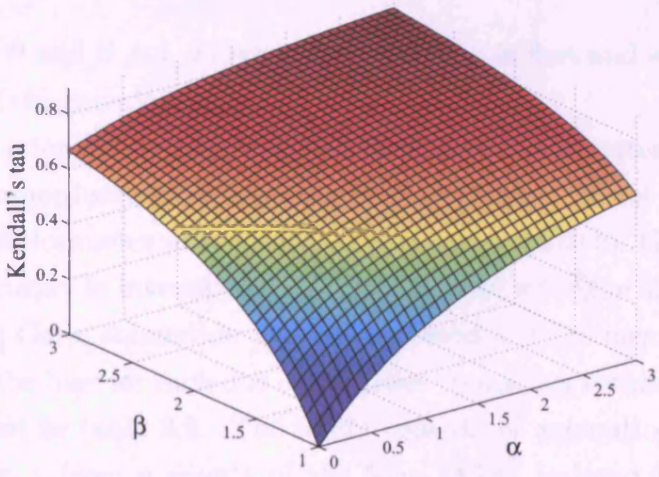


Figure 3.5: Kendall's tau for the BB1 copula.

The estimators (3.7) and(3.8) take now the form:

$$(\alpha^*, \beta^*) = \arg\min_{\alpha,\beta} \left\{ \sum_{i=1}^{n} \left(z + \frac{z^{\alpha+1}(z^{-\alpha}-1)}{\alpha\beta} - K_n(z_i) \right)^2 \right\}^{1/2}, \tag{3.18}$$

and

$$(\hat{\alpha}, \hat{\beta}) = \arg\max_{\alpha,\beta} \sum_{i=1}^{n} \log\left(1 + \frac{1}{\alpha\beta} - \frac{1+\alpha}{\alpha\beta} z_i^{\alpha}\right). \tag{3.19}$$

The competitor of the last two estimators is the method-of-moment which is considered to be the standard method for the estimation of an Archimedean copula in cases where the margins F, G of the associated joint distribution $H(x, y) = C(F(x), G(y))$, remain unspecified. For single-parameter families, the estimation involves a measure of association, in most cases Kendall's tau. A comparison of the estimators was carried out for the Frank family in table 3.1. When a vector parameter is under consideration the extension of the method-of-moments involves the first moments of the distribution K. For a copula of the type (3.13) it is straightforward to show that

$$\mathbb{E}(Z) = \frac{(\alpha+2)\beta - 1}{2\beta(\alpha+2)} \quad \text{and} \quad \mathbb{E}(Z^2) = \frac{(\alpha+3)\beta - 2}{3\beta(\alpha+3)}$$

Hence, the method of moments estimator is

$$(\tilde{\alpha}, \tilde{\beta}) = \arg\min_{\alpha,\beta} [(\mathbb{E}(Z) - m_1)^2 + (\mathbb{E}(Z^2) - m_2)^2] \tag{3.20}$$

subject to $\alpha > 0$ and $\beta \geq 1$, where m_1, m_2 the sample first and second moments around zero of the pseudosample.

The solution for (3.18), (3.19) or (3.20), is obtained in terms of a numerical optimization procedure, since an analytical expression is not possible. The small sample performance of the three competing methods for the estimation of a vector parameter, is investigated in a similar way with the single parameter case. A Monte Carlo simulation was implemented and the mean squared error as well as the the bias for each one of the three competing estimators, $\alpha^*, \hat{\alpha}$ and $\tilde{\alpha}$, was recorded in table 3.2. The study consists of generating 1000 random samples of size n from a copula of the form (3.13), indexed by (α, β). The sample size takes the values, $n = 15, 25$ or 50, while the parameters α and β take values from the set $\{1, 2, 3\}$, covering a wide range for Kendall's tau; from $\tau = 0.33$ for $(\alpha, \beta) = (1, 1)$ to $\tau = 0.86$ when $(\alpha, \beta) = (3, 3)$.

The most important results of the study are:

α	β		Bias (× 100) $n=15$		$n=25$		$n=50$		Mean squared error (× 10) $n=15$		$n=25$		$n=50$	
1	1	α^*	-17	13	-24	14	-20	10	6	0	3	0	2	0
		$\tilde{\alpha}$	89	22	31	15	4	8	34	3	7	1	2	0
		$\hat{\alpha}$	92	22	45	14	17	7	29	2	8	0	2	0
1	2	α^*	62	-21	33	-13	24	-9	37	10	18	6	11	3
		$\tilde{\alpha}$	385	168	158	63	62	17	300	126	93	31	19	6
		$\hat{\alpha}$	182	186	117	65	64	17	112	120	64	22	25	4
1	3	α^*	113	-62	72	-52	31	-18	80	30	43	16	19	10
		$\tilde{\alpha}$	576	307	287	134	83	67	489	226	219	100	41	31
		$\hat{\alpha}$	199	335	142	154	72	55	123	228	84	85	33	18
2	1	α^*	-53	23	-47	18	-37	13	16	2	11	1	6	0
		$\tilde{\alpha}$	161	50	64	19	13	11	97	18	27	1	6	0
		$\hat{\alpha}$	107	55	54	23	21	10	67	16	25	2	8	0
2	2	α^*	61	-8	26	-2	28	1	88	15	41	8	29	6
		$\tilde{\alpha}$	494	251	302	89	147	24	388	186	226	57	80	12
		$\hat{\alpha}$	172	281	112	131	88	41	139	196	92	57	60	10
2	3	α^*	155	-27	100	-30	42	-11	208	55	119	28	52	17
		$\tilde{\alpha}$	690	429	465	204	200	70	547	307	359	151	140	49
		$\hat{\alpha}$	194	441	111	269	68	110	130	309	83	183	56	42
3	1	α^*	-88	29	-77	25	-64	19	38	3	24	2	14	1
		$\tilde{\alpha}$	228	79	110	30	20	15	154	40	65	6	13	1
		$\hat{\alpha}$	99	88	54	39	23	16	88	39	49	6	21	1
3	2	α^*	79	2	48	-1	4	9	155	26	79	15	46	9
		$\tilde{\alpha}$	538	356	408	115	217	30	384	265	292	81	142	18
		$\hat{\alpha}$	134	369	98	193	51	73	127	272	106	110	70	20
3	3	α^*	191	-7	87	-5	67	-6	436	81	204	44	108	23
		$\tilde{\alpha}$	649	517	531	291	326	64	455	364	374	197	225	52
		$\hat{\alpha}$	169	487	58	396	53	165	129	347	76	254	65	72

Table 3.2: Mean squared error (MSE) from a Monte Carlo study for the three estimators. (α^*, β^*): the minimum distance estimator; $(\tilde{\alpha}, \tilde{\beta})$: the method of moments estimator; $(\hat{\alpha}, \hat{\beta})$: maximum likelihood estimator.

- The estimation based on (α^*, β^*), seems to perform better in terms of the mean squared error, than its competitors $(\tilde{\alpha}, \tilde{\beta})$, $(\hat{\alpha}, \hat{\beta})$, for all sample sizes and all values for the parameters α, β.

- Estimator (3.18) tends to underestimate one of the parameters. For relatively low values of Kendall's tau, the approach gives negative bias for α and possitive for β. The reverse behaviour is observed as Kendall's tau increases.

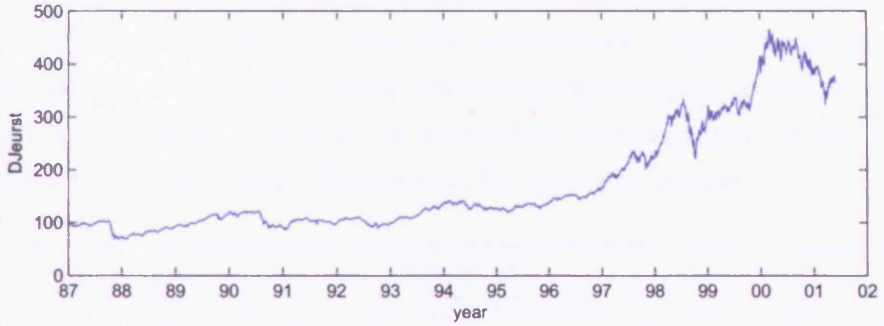
- For large values of Kendall's tau, the method of moments tends to achieve smaller MSE than the maximum likelihood approach for the estimation of β. This does not however change the overall picture for the performance of the three methods that points out (α^*, β^*) as the most well behaved estimator of the three examined in the study.

3.3 Real data application.

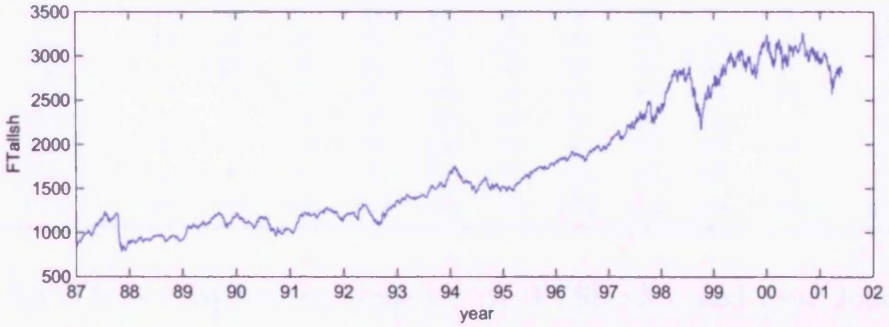
This section illustrates the methods developed in this chapter and presents the results of the analysis of a large bivariate dataset. The data consist of the daily returns of two financial indices, namely the DJ EURO STOXX and the FTSE ALL SHARE, whose values were gathered over the period from January 1st, 1987 up to May 31st 2001 to form a bivariate sample of 3760 observations. A plot of the raw indices' values is drawn in figure 3.6 and a scatter plot of the returns is included in figure 3.7.

The aim of the study is to investigate which family of Archimedean copulas gives the best fit for the dataset and if possible to identify the underlying generating law. The choice is between four copulas; the widely used one-parameter Clayton, Gumbel and Frank families and the two-parameter one of the form (3.13).

Table 3.3 is a 10×10 cross classification of the two variables of interest, X and Y, the daily percentage returns for the FTSE and Dow-Jones indices respectively. Each observation (x, y) was ordered componentwise and the cells' upper boundaries for the cross classification table were chosen as the order statistics $X_{(k)}$ or $Y_{(k)}$ of rank $[3760\frac{k}{10}]$ for $k = 1, \ldots, 10$, where $[x]$ denotes the rounded to the closest integer of x. Each entry in table 3.3 represents the number



(a) Dow Jones daily data



(b) FT daily data.

Figure 3.6: Plot of the two indices, Dow Jones and FT, for the period between January 1st, 1987 and May 31st 2001. For each index, the total number of daily observations for this period, amounts to 3761.

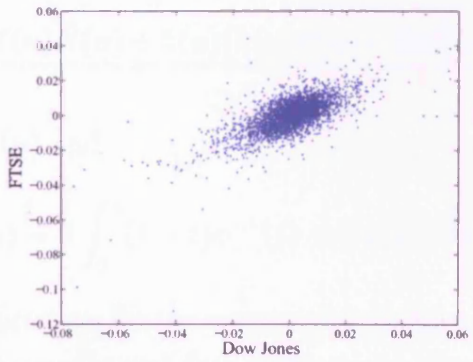


Figure 3.7: Scatter plot of the daily percentage returns for Dow Jones and FTSE.

of the ranked observations componentwise less than or equal to the cell's upper boundaries and strictly greater than the previous cell's upper boundaries. The column and the row labelled 'total' represent the marginal densities which as expected are uniform. Table 3.3 is treated as representative of the copula's density for the returns' joint distribution for the 10×10 grid.

	Upper Bounds										
Upper Bounds	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$	$X_{(8)}$	$X_{(9)}$	$X_{(10)}$	Total
$Y_{(1)}$	196	81	36	26	6	7	7	6	6	5	376
$Y_{(2)}$	74	77	70	51	27	25	21	12	14	5	376
$Y_{(3)}$	34	67	69	51	41	41	24	17	22	10	376
$Y_{(4)}$	21	50	46	66	47	48	36	30	23	9	376
$Y_{(5)}$	14	26	50	46	67	59	36	35	29	14	376
$Y_{(6)}$	9	26	36	54	48	48	56	47	33	19	376
$Y_{(7)}$	12	19	24	35	48	49	60	49	47	33	376
$Y_{(8)}$	10	17	25	27	49	44	56	67	46	35	376
$Y_{(9)}$	4	7	14	13	29	38	53	60	89	69	376
$Y_{(10)}$	2	6	6	7	14	17	27	53	67	177	376
Total	376	376	376	376	376	376	376	376	376	376	3760

Table 3.3: Cross classification table for the FTSE (X), and Dow Jones (Y), dialy returns.

It was established that an Archimedean copula C is characterized by the distribution $K(z) = z - \frac{\phi(z)}{\phi'(z)}$ of the random variable $Z = C(U, V)$. The empirical distribution K_n of the pseudosample yield by (3.3) is regarded as an empirical estimate of K. Genest and Rivest (1993) have shown that an $o(\frac{1}{n})$ approximation of the asymptotic variance of K_n is given by:

$$\frac{K(u)\bar{K}(u) + k(u)\big(k(u)R(u) - 2u\bar{K}(u)\big)}{n} \tag{3.21}$$

where $\bar{K}(u) = 1 - K(u)$ and

$$R(u) = 2\int_0^1 (1-t)\phi^{-1}\big((1+t)\phi(u)\big)\,\mathrm{d}t - u^2$$

Unfortunately the expression for the asymptotic variance depends on K which is usually treated as unknown in real data applications. Even if a specific parametric form is assumed for K, (3.21) does not usually have a closed analytical form. Among the four systems of Archimedean copulas considered in this section, only

the Clayton family yields an explicit expression:

$$R(u) = \frac{2\alpha u\big(-\alpha + (1-2\alpha)(1-u^\alpha) + \alpha(2-u^\alpha)^{2-\frac{1}{\alpha}}\big)}{(1-2\alpha)(1-\alpha)(1-u^\alpha)^2} - u^2$$

and $K(u) = u(\frac{1-u^\alpha}{\alpha} + 1)$

In the following we adopt Genest and Rivest (1993) and obtain the approximate asymptotic variance $\sigma^2_{\tilde{\alpha}}$ of the estimator K_n from (3.21) for K, R and $\tilde{\alpha}$, the method of moments estimate of α, all corresponding to the Clayton case.

Because the study concerns the identification of the most proper family rather than the selection of a member from a specific class we draw familywise 95% level confidence bands for $K_n = K_n(z)$ from $K_n(z) \pm 4.72\,\sigma_{\tilde{\alpha}}(z)$. The value 4.72 was proposed by Kotel'nikova and Chmaladze (1982) and was used in Genest and Rivest (1993). It serves exactly the purpose of 'across-families' confidence bands rather than the traditional choice of 1.96 that concerns 'within-the-family' confidence levels. Figure 3.8 depicts the empirical estimate $z - K_n(z)$ along with the confidence bands which were calculated in the manner described previously. Overimposed in the same graph, are included the estimated $z - K(z)$

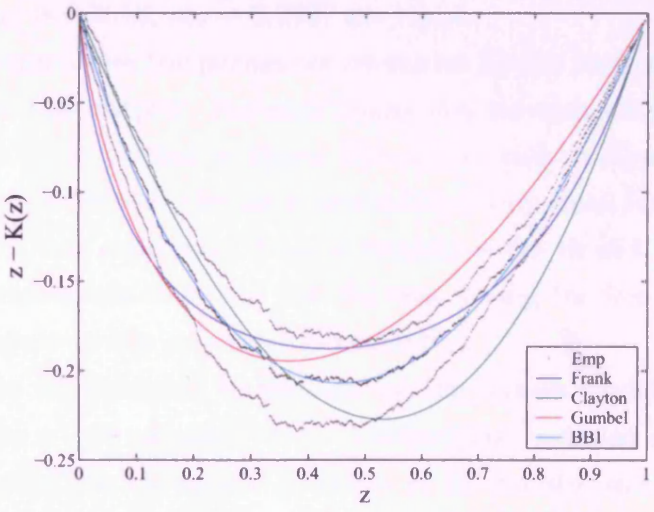


Figure 3.8: Plot of the empirical estimate K_n and its 95% confidence bands. Overimposed, are parametric estimates from four different families, where the parameters are α^*, β^*. —, Frank; —, Clayton; —, Gumbel; —, BB1 copula.

Family	$\hat{\alpha}$	$\hat{\beta}$	$\tilde{\alpha}$	$\tilde{\beta}$	α^*	β^*	χ^2
Clayton	1.541	-	1.604	-	1.331	-	372.689
Gumbel	1.817	-	1.802	-	1.890	-	283.364
Frank	4.595	-	4.819	-	4.608	-	263.932
BB1	0.587	1.409	0.568	1.408	0.593	1.383	88.109

Table 3.4: Estimated parameters for four copula families. $\hat{\alpha}$ and $\tilde{\alpha}$ are the MLE or method-of-moments on the pseudobservations respectively and α^*, β^* is the minimum distance estimator. For α^*, β^* we include the chi-square statistic to assess the goodness of fit.

for the four different families concerned herein. The parameters α or β have been estimated by α^* and β^*.

The sample value of Kendalls tau for the dataset is $\tau = 0.4452$. Using the identity $\tau = 4\mathbb{E}(Z) - 1$, for each one of three one-parameter fitted families we derive the method-of-moments estimates in such a manner that the theoretical value matches the observed one. For the two-parameter family we have to construct the pseudosample $Z_1, Z_2, \ldots, Z_n$ and then obtain the estimates $\tilde{\alpha}$, $\tilde{\beta}$ as the values for which the theoretical moments $\mathbb{E}(Z)$, $\mathbb{E}(Z^2)$ and the sample counterparts $m_1 = 0.3618$, $m_2 = 0.2007$ are equal.

Table 3.4 summarizes the parameter estimates for the four competing methods discussed in this chapter. For each family and for each estimate, the corresponding $z - K(z)$ is plotted in figure 3.9. A first look at those results shows that the differences between the three methods are very small for large datasets, something which was expected. What is striking is the fit of the family (3.13) to the data, which seems to be not just the best among the families tested here but also quite close to the empirical estimate too.

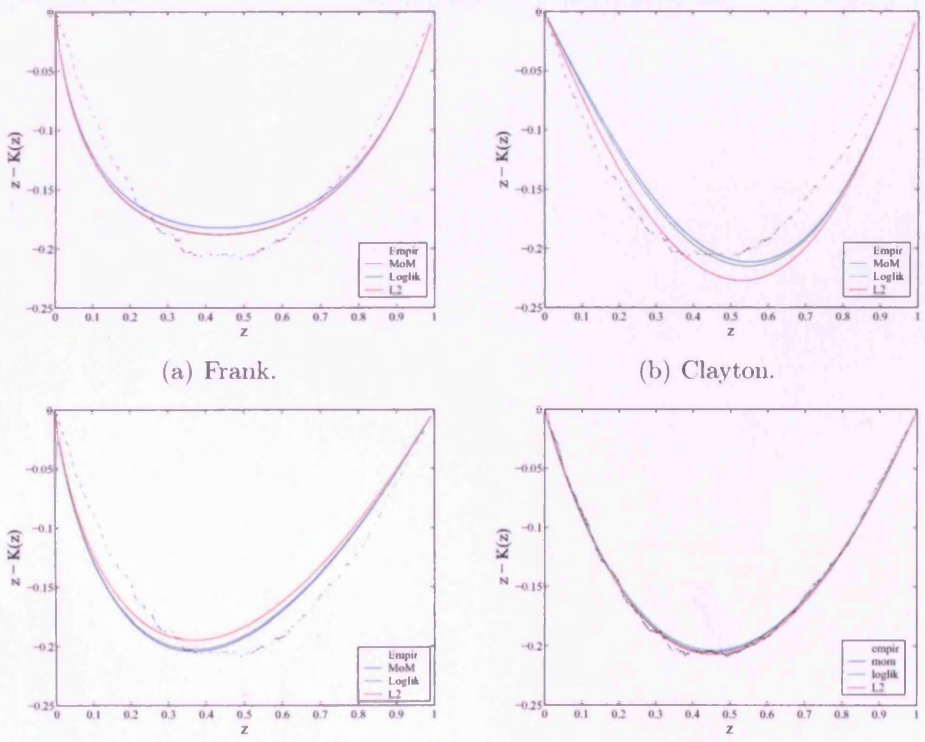
To assess the fit, predicted frequencies for the various models as those were implied from the α^*, β^* estimator, were obtained and included in figure 3.10.

For each model the chi-square goodness of fit statistic was computed and its values were recorder in the last column in table 3.4. Those results align to the conclusions drawn from the figures and indicate that the model of the form (3.13) is the most appropriate for the specific dataset. The most important result however, is that the chi-square statistic does not give evidence against the hypothesis that the data have a copula of the form (3.13) at 95% level. In

fact, $1 - \chi^2_{81}(88.109) = 0.27$. It seems therefore, that the family

$$C(u,v) = \left[1 + \left((u^{-\alpha} - 1)^{\beta} + (v^{-\alpha} - 1)^{\beta}\right)^{1/\beta}\right]^{-1/\alpha} \qquad \alpha > 0,\ \beta \geq 1$$

arises as a reasonable choice for the modelling of the joint distribution of positively correlated financial returns, something, which to the knowledge of the author, has not been mentioned before by researchers in the field. A closer look at the table 3.3 of the observed frequencies shows that the joint density of the ranks of the random variables presents clustering close to the origin (lower left quadrant) and to point (1,1) , (upper right quadrant). The effect however, is stronger to the origin than to the upper tail. There also seems to be a symmetric pattern across the line jointing those two points. It appears from figure 3.10 of



(a) Frank. (b) Clayton.

(c) Gumbel. (d) BB1.

Figure 3.9: Fit of four different families. Four each family we illustrate the empirical estimate (Empir.), and the parametric estimates using the method of moments (MoM), maximum likehood (LogLik) or the minimum distance (L_2) approach.

the estimated frequencies, that the latter is accommodated by all models tested here. Each family however, assumes different patterns of clustering at the tails. Frank copula for example, implies a symmetric density with respect to both diagonals, whereas the Clayton and Gumbel families imply that observations accumulate on the two quadrants in an asymmetric manner. Specifically, the Clayton class produces a more severe to the lower tail than to the lower one where the effect is milder. On the other way, the Gumbel class has almost the reverse effect by assuming a severe clustering to the upper tail and a weaker to the lower one. The copula of the form (3.13) seems to be more flexible that its competitors by exhibiting both upper and lower tail dependence as already shown in (3.14) and (3.15). A look at the table of the observed frequencies reveal that the clustering effect, is present at both tails and seems to be slightly stronger on the lower tail, something which is accommodated by family (3.13).

	Upper Bounds									
Upper Bounds	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$	$X_{(8)}$	$X_{(9)}$	$X_{(10)}$
$Y_{(1)}$	227	73	31	16	10	7	4	4	2	2
$Y_{(2)}$	73	95	65	45	30	21	17	12	10	8
$Y_{(3)}$	31	65	66	53	43	35	27	22	19	15
$Y_{(4)}$	16	45	53	53	49	42	37	31	27	23
$Y_{(5)}$	10	30	43	49	48	46	43	40	35	32
$Y_{(6)}$	7	21	35	42	46	48	47	45	44	41
$Y_{(7)}$	4	17	27	37	43	47	50	51	50	50
$Y_{(8)}$	4	12	22	31	40	45	51	54	58	59
$Y_{(9)}$	2	10	19	27	35	44	50	58	62	69
$Y_{(10)}$	2	8	15	23	32	41	50	59	69	77

(a) Clayton.

	Upper Bounds									
Upper Bounds	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$	$X_{(8)}$	$X_{(9)}$	$X_{(10)}$
$Y_{(1)}$	136	79	54	37	26	19	12	8	4	1
$Y_{(2)}$	79	75	62	50	39	28	20	13	7	3
$Y_{(3)}$	54	62	61	55	47	37	28	18	11	3
$Y_{(4)}$	37	50	55	56	52	45	35	26	15	5
$Y_{(5)}$	26	39	47	52	53	52	45	34	21	7
$Y_{(6)}$	19	28	37	45	52	54	54	46	30	11
$Y_{(7)}$	12	20	28	35	45	54	60	59	45	18
$Y_{(8)}$	8	13	18	26	34	46	59	70	70	32
$Y_{(9)}$	4	7	11	15	21	30	45	70	99	74
$Y_{(10)}$	1	3	3	5	7	11	18	32	74	222

(b) Gumbel.

	Upper Bounds									
Upper Bounds	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$	$X_{(8)}$	$X_{(9)}$	$X_{(10)}$
$Y_{(1)}$	121	87	59	40	27	17	11	7	4	3
$Y_{(2)}$	87	77	65	49	35	24	17	10	8	4
$Y_{(3)}$	59	65	62	55	44	34	23	17	10	7
$Y_{(4)}$	40	49	55	56	51	42	32	23	17	11
$Y_{(5)}$	27	35	44	51	52	50	42	34	24	17
$Y_{(6)}$	17	24	34	42	50	52	51	44	35	27
$Y_{(7)}$	11	17	23	32	42	51	56	55	49	40
$Y_{(8)}$	7	10	17	23	34	44	55	62	65	59
$Y_{(9)}$	4	8	10	17	24	35	49	65	77	87
$Y_{(10)}$	3	4	7	11	17	27	40	59	87	121

(c) Frank.

	Upper Bounds									
Upper Bounds	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$	$X_{(8)}$	$X_{(9)}$	$X_{(10)}$
$Y_{(1)}$	193	75	40	24	16	10	8	5	3	2
$Y_{(2)}$	75	83	64	47	34	27	18	14	9	5
$Y_{(3)}$	40	64	62	54	46	36	29	21	16	8
$Y_{(4)}$	24	47	54	56	50	44	38	30	21	12
$Y_{(5)}$	16	34	46	50	52	50	44	38	30	16
$Y_{(6)}$	10	27	36	44	50	51	51	47	37	23
$Y_{(7)}$	8	18	29	38	44	51	54	54	49	31
$Y_{(8)}$	5	14	21	30	38	47	54	60	63	44
$Y_{(9)}$	3	9	16	21	30	37	49	63	76	72
$Y_{(10)}$	2	5	8	12	16	23	31	44	72	163

(d) BB1.

Figure 3.10: Tables of expected frequencies for the four families.

Chapter 4

Non-parametric Bayesian density estimation.

4.1 Introduction

In this section we present a Bayesian non-parametric approach for the estimation of the joint density c of a copula C. Although the method is applicable to any copula in general, we apply it to members of the Archimedean class. Illustrative examples of the approach are included towards the end of this chapter, in section 4.6. The work was inspired by similar investigations by Heikkinen and Arjas (1998) and Green (1995). Briefly, the true joint density c is represented by a random piecewise constant function on the unit square. We put a prior distribution on the piecewise constant function which is updated with information from the observed data and derive the posterior distribution. The aim is to obtain a sample from that posterior distribution; a sample consisting of piecewise functions with different characteristics, since by design both the jump locations and the corresponding attained values vary between different functions. Finally, we approximate c, with the mean from the posterior sample. Note that although the sample consists of piecewise constant functions, the mean, in the limit and as the sample size tends to infinity, will be a smooth function.

The approach which is introduced in this chapter lies in the wide computational field of Markov Chain Monte Carlo (MCMC) methods and as such it generates a dependent sample. By construction however, the kernel of the chain has the desired distribution as its invariant distribution. In these cases, Markov

chain theory ensures that the mean of the dependent MCMC sample converges to the mean of the invariant distribution. A natural 'point' estimate for the theoretical density is simply the sample mean.

The remainder of this chapter is organized as follows. In the next two section we present the basic elements of the MCMC theory and then we proceed to section (4.4) where we specify the probabilistic features of the model. The details of the sampler used to obtain the approximate sample from the target distribution are discussed in section (4.5). The methodology is illustrated in section (4.6) with applications based on simulated data from Archimedean copulas, but also on a real world financial dataset.

Before proceeding to the next section, we should underline a key difference with the methods presented in the previous chapter where the margins F, G of the bivariate distribution $H(x, y) = C\big(F(x), G(y)\big)$ remained unspecified. Here, the margins are considered given. For the real data illustrations included in the last sections of the present chapter we use empirical estimates for F and G to transform the observations from $\mathbb{R}^2$ to a sample on the unit square drawn from the unknown C.

Note that despite the use of piecewise constant functions to describe the unknown density, the method is considered as non-parametric, because the sampling, by design, is done in an infinite dimensional space where the data drive the inference. A short revision of the MCMC theory now follows.

4.2 Standard Markov Chain Monte Carlo.

Many statistical problems involve the calculation of an integral of the form:

$$\mathbb{E}_\pi(h(X)) = \int_{\mathbb{R}^d} h(x)\pi(\mathrm{d}x) \tag{4.1}$$

where π, defined on $\mathbb{R}^d$, is the d-dimensional distribution of the random variable X and h some function of X. If we can draw n independent random variables $X_1, X_2, \ldots, X_n$ from π, then the Strong Law of Large Numbers implies

$$\frac{\sum_i^n h(X_i)}{n} \to \mathbb{E}_\pi(h(X)) \quad a.s.$$

as $n \to \infty$ and we obtain an estimate of the population mean. However independent sampling is not always feasible. MCMC theory provides a framework

to obtain an approximate dependent sample from π and bypass the integration in equation 4.1, when this is analytically intractable.

It is common in Markov Chain theory to start from a given transition kernel $K(x, A)$ and try to derive the invariant distribution π:

$$\pi(A) = \int_{\mathbb{R}^d} K(x, A)\pi(\mathrm{d}x)$$

The transition kernel $K(x, A)$ is a function of a point $x \in \mathbb{R}^d$ and a Borel set $A \subset \mathbb{R}^d$. For fixed x, the kernel is a distribution on $\mathbb{R}^d$ and it denotes the conditional distribution of 'moving from x somewhere into A'. Transitions from x to x are also allowed, hence $K(x, \{x\})$ is not necessary zero.

MCMC methods are concerned with the reverse problem. The density of the invariant distribution is known (it is the distribution π from which samples are needed). What we want to find is the transition kernel. That is be used to construct a Markov Chain $X_1, X_2, \ldots$ for which, under some rather general conditions, the ergodic theorem says:

$$\frac{\sum_i^n h(X_i)}{n} \rightarrow \mathbb{E}_\pi(h(X)) \quad a.s.$$

Although the Markov chain is not an i.i.d. sample from π, the ergodic theorem guarantees that asymptotically, the sample mean tends to the population mean. This is the object of Markov Chain Monte Carlo (MCMC) methods, a term which is used for any approach that generates an ergodic Markov Chain whose invariant distribution is π.

Let's now decompose the measure induced by K into the sum of a two measures:

$$K(x, A) = \mu_x(A) + r(x)\delta_x(A)$$

where $\delta_x(A)$ the Dirac measure of A. When we are at point x, suppose we propose a move to set $\mathrm{d}y$. Express now μ as:

$$\mu_x(A) = \int_A p(x, \mathrm{d}y)\alpha(x, y),$$

where $p(x, \mathrm{d}y)$ is some measure on $\mathbb{R}^d$ indexed by x (assumed given) and α is the probability of the proposed move $x \rightarrow \mathrm{d}y$ being accepted. More clearly, for fixed x and $\mathrm{d}y$, $\alpha : \alpha(x, y)$ is a probability measure on a space consisted by two

disjoint sets: accept and do not accept the proposed move. Note also that p is a subprobability measure on $\mathbb{R}^d$ indexed by x, hence $p(x, \mathbb{R}^d) < 1$. For point x, we do not propose a change with probability $1 - p(x, \mathbb{R}^d)$ and with probability $\int_{\mathbb{R}} p(x, \mathrm{d}y)(1 - \alpha(x, y))$ the proposed change is not accepted. Therefore, the probability of not moving from x is:

$$\begin{aligned} r(x) &= \int_{\mathbb{R}^d} p(x, \mathrm{d}y)(1 - \alpha(x, y)) + (1 - p(x, \mathbb{R}^d)) \\ &= 1 - \int_{\mathbb{R}^d} p(x, \mathrm{d}y)\alpha(x, y). \end{aligned}$$

The transition kernel can now be rewritten:

$$K(x, A) = \int_A p(x, \mathrm{d}y)\alpha(x, y) + r(x)\delta_x(A).$$

Then:

$$\begin{aligned} \int_{\mathbb{R}^d} K(x, A)\pi(\mathrm{d}x) &= \int_{\mathbb{R}^d} \Big(\int_A p(x, \mathrm{d}y)\alpha(x, y) + r(x)\delta_x(A) \Big) \pi'(x)\mathrm{d}x \\ &= \int_{\mathbb{R}^d} \int_A p(x, \mathrm{d}y)\pi'(x)\alpha(x, y)\mathrm{d}x + \int_{\mathbb{R}^d} r(x)\delta_x(A)\pi'(x)\mathrm{d}x \\ &= \int_{\mathbb{R}^d} \pi'(x) \int_A p(x, \mathrm{d}y)\alpha(x, y)\mathrm{d}x + \int_{\mathbb{R}^d} r(x)\delta_x(A)\pi'(x)\mathrm{d}x. \end{aligned} \tag{4.2}$$

If we assume that the detailed balance condition:

$$\int_{\mathbb{R}^d} \pi'(x) \int_A p(x, \mathrm{d}y)\alpha(x, y)\mathrm{d}x = \int_A \pi'(y) \int_{\mathbb{R}^d} p(y, \mathrm{d}x)\alpha(y, x)\mathrm{d}y, \tag{4.3}$$

is satisfied, equation(4.2) becomes:

$$\begin{aligned} \int_{\mathbb{R}^d} K(x, A)\pi(\mathrm{d}x) &= \int_A \pi'(y) \int_{\mathbb{R}^d} p(y, \mathrm{d}x)\alpha(y, x)\mathrm{d}y + \int_{\mathbb{R}^d} r(x)\delta_x(A)\pi'(x)\mathrm{d}x = \\ &= \int_A \pi'(y)(1 - r(y))\mathrm{d}y + \int_{\mathbb{R}^d} r(x)\delta_x(A)\pi'(x)\mathrm{d}x = \\ &= \int_A \pi'(y)\mathrm{d}y - \int_A r(y)\pi'(y)\mathrm{d}y + \int_A r(x)\pi'(x)\mathrm{d}x = \\ &= \int_A \pi'(y)\mathrm{d}y, \end{aligned}$$

and hence π is the invariant distribution of the transition kernel K. Assumption (4.3) is also known as the reversibility condition because the LHS expresses the

probability of making and accepting a move from $\mathbb{R}^d \rightarrow A$, while RHS is the probability of making and accepting the reverse move A$\rightarrow \mathbb{R}^d$. Although this is a sufficient and not necessary condition, it is common to impose it because it facilitates the task of constructing an appropriate transition kernel. Suppose now that $\pi(\mathrm{d}x)p(x,\mathrm{d}y)$ has density $f(x,y)$ with respect to the Lebesque measure on $\mathbb{R}^d \times \mathbb{R}^d$. Then, for all Borel sets A, B,

$$\int_A \pi(\mathrm{d}x) \int_B p(x,\mathrm{d}y)\alpha(x,y) = \int_A \int_B \lambda(\mathrm{d}x,\mathrm{d}y) f(x,y)\alpha(x,y),$$

and under the assumption:

$$f(x,y)\alpha(x,y) = f(y,x)\alpha(y,x), \tag{4.4}$$

we derive:

$$\begin{aligned}
\int_A \pi(\mathrm{d}x) \int_B p(x,\mathrm{d}y)\alpha(x,y) &= \int_A \int_B \lambda(\mathrm{d}x,\mathrm{d}y) f(x,y)\alpha(x,y) \\
&= \int_B \int_A \lambda(\mathrm{d}y,\mathrm{d}x) f(x,y)\alpha(x,y) \quad \text{(by symmetry of Lebesque)} \\
&= \int_B \int_A \lambda(\mathrm{d}y,\mathrm{d}x) f(y,x)\alpha(y,x) \quad \text{(by 4.4)} \\
&= \int_B \pi(\mathrm{d}y) \int_A p(y,\mathrm{d}x)\alpha(y,x)
\end{aligned}$$

Therefore, the detailed balance assumption in (4.3) holds only for an appropriate α as (4.4) demonstrates. A suitable choice is:

$$\alpha(x,y) = \min\left\{1, \frac{f(y,x)}{f(x,y)}\right\}. \tag{4.5}$$

Indeed, if $\alpha(x,y) = \frac{f(y,x)}{f(x,y)} < 1$ then $\alpha(y,x) = \min\left\{1, \frac{f(x,y)}{f(y,x)}\right\} = 1$ because $\frac{f(x,y)}{f(y,x)} = \frac{1}{\alpha(x,y)} > 1$ and relation (4.4) is true. Also, it proves to be more meaningful to express the joint density of the random vector (X,Y) as the product $\pi(\mathrm{d}x)p(x,\mathrm{d}y)$ and to rewrite the acceptance probabilities as:

$$\alpha(x,y) = \min\left\{1, \frac{p(y,\mathrm{d}x)\pi'(y)}{p(x,\mathrm{d}y)\pi'(x)}\right\}. \tag{4.6}$$

It follows now that the transition kernel,

$$K(x,A) = \int_A p(x,\mathrm{d}y)\alpha(x,y) + \int_{\mathbb{R}^d} p(x,\mathrm{d}y)(1-\alpha(x,y)) + (1-p(x,\mathbb{R}^d)), \tag{4.7}$$

where α as in (4.6), satisfies by construction assumption (4.3) and hence, it has π as its invariant distribution. The issue now is to sample from (4.7). Term $\int_A p(x, \mathrm{d}y)\alpha(x, y)$ expresses the joint probability of making a move from x to A and accept the move (for $x \notin A$). Conditionally on x, a clearer notation may be:

$$\begin{aligned}
&\mathbb{P}\{(\text{propose move } x \to A), (\text{accept move } x \to A)\} \\
&= \int_A p(\text{prop. move } x \to \mathrm{d}y)\, p\{(\text{accept move } x \to \mathrm{d}y) \,|\, (\text{prop. move } x \to \mathrm{d}y)\} \\
&= \int_A p(x, \mathrm{d}y)\alpha(x, y).
\end{aligned}$$

Consider now the case $A = \{x\}$ and take for simplicity $\int_{\mathbb{R}^d} p(x, \mathrm{d}y) = 1$ (which is usually the case). Then, equation (4.7) becomes:

$$K(x, A) = \int_{\mathbb{R}^d} p(x, \mathrm{d}y)(1 - \alpha(x, y)),$$

the joint probability of proposing a move anywhere in the domain and reject the move (alternatively, it is the expectation of the point function $\alpha(x, y)$ with respect to $1 - p(x, \mathrm{d}y)$). Therefore, to sample from (4.7) we follow the following template, introduced by Metropolis et al. (1953) and Hastings (1970) and is widely known as the Metropolis-Hastings (MH) algorithm:

Given an intitial value x_0

- For $i = 0, 1 \ldots n$:
- generate $Y_i \sim p(x_i, \mathrm{d}y)$ and U from $U(0, 1)$
- if $u \leq \alpha(x_i, y_i)$ set $x_{i+1} = y_i$
- else set $x_{i+1} = x_i$
- return the sample $(x_1, x_2, \ldots, x_{i+1})$

Because asymptotically, $K_n(\cdot, A) \to \pi(\cdot)$, point x_n can be considered to be drawn from π, and from the ergodic theorem, $\bar{X}_n \to \mathbb{E}_\pi(X)$.

4.3 Reversible jump Markov Chain Monte Carlo.

Green (1995) expands MCMC methodology to cover a more general space of the form $\mathcal{C} = \bigcup_k \{k\} \times \mathbb{R}^{d_k}$ and introduces moves between (sub)spaces of different dimensionality. Each pair of the one-way moves that take us from a point $x \in \mathbb{R}^{d_1}$ to a set $dy \subset \mathbb{R}^{d_2}$ and then back to $\mathbb{R}^{d_1}, d_1 \neq d_2$ is denoted by m. Then the transition kernel is:

$$K(x, A) = \sum_m \int_A p_m(x, dy)\alpha(x, y) + r(x)\delta_x(A),$$

for

$$r(x) = \int_{\mathcal{C}} p_m(x, dy)(1 - \alpha(x, y)) + \Big(1 - \sum_m p_m(x, \mathcal{C})\Big).$$

Following almost the same pattern as in the standard MCMC we must guarantee that for all $(x \times y) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$,

$$f^{\xi_m}(x \times y)\alpha(x, y) = f^{\xi_m}(y \times x)\alpha(y, x), \tag{4.8}$$

where f^{ξ_m} is the density of $\pi(dx)p_m(x, dy)$ with respect to a symmetric measure ξ_m on $\mathcal{C} \times \mathcal{C}$. The trick is to 'project' the product space $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ onto some properly chosen $\mathbb{R}^d$. Suppose, without loss of generality, that $d_1 < d_2$. We complement x with a random, from an arbitrary distribution, vector $u_1 \in \mathbb{R}^{m_1}$, such as $(x, u_1) \in \mathbb{R}^{d_1+m_1} = \mathbb{R}^{d_2}$. The move across the two spaces is done via a function $g : g(x, u_1) = y$ which, given knowledge of u_1, is of deterministic nature and up to the modeler to choose. Hence, being at x, the point y where we 'land', is determined by u_1. It is worth mentioning that the length m_1 of u_1 is chosen such as $d_1 + m_1 = d_2$, because g must be a bijection and hence, the dimesion of the domain should match that of the range. We can now construct a symmetric measure ξ_m in terms of the Lebesque on $\mathbb{R}^{d_1+m_1}$. Specifically, for all Borel sets A, B

$$\begin{aligned}\xi_m(A \times B) &= \lambda\{(x, u_1) : x \in A, u_1 : y = g(x, u_1) \in B\} \\ &= \lambda\{y : y \in B, x = g^*(y) \in A\} = \xi(B \times A),\end{aligned}$$

where g^* is another function chosen at will and describes the reverse move. The gain, apart from symmetry, is that the 'reference' space is now some multidimensional real space with the corresponding user-friendly Lebesque measure. Then we define the density f^{ξ_m} of the joint distribution $\pi(A)p_m(x, B)$ with respect to the ξ_m measure, in terms of the density f^{λ} with respect to λ. More formally, if $j(x)$ is the probability of proposing the cross space move form $x \in A$ to $y \in B$,

$$\begin{aligned} f^{\xi_m}(x \times y) &= f^{\lambda}_{X,U_1}(x, u_1)j(x) = p(x)q(u_1 \mid x)j(x), \\ f^{\xi_m}(y \times x) = \; f^{\lambda}_{Y}(y)j(y) &= f^{\lambda}_{X,U_1}(g(x, u_1)) \left|\tfrac{\partial(g(x,u_1))}{\partial(x,u_1)}\right| j(y) \\ &= p(g(x, u_1)) \left|\tfrac{\partial(g(x,u_1))}{\partial(x,u_1)}\right| j(y), \end{aligned} \tag{4.9}$$

and requirement (4.8) is satisfied by setting

$$\begin{aligned} \alpha(x, y) &= \frac{f^{\xi_m}(y \times x)}{f^{\xi_m}(x \times y)} \\ &= \frac{p(g(x, u_1))j(y)}{p(x)q(u_1 \mid x)j(x)} \left|\frac{\partial(g(x, u_1))}{\partial(x, u_1)}\right| \end{aligned}$$

When data enter the computation the last relation becomes:

$$\alpha(x, y) = \frac{p(g(x, u_1) \mid \text{data})j(y)}{p(x \mid \text{data})q(u_1 \mid x, \text{data})j(x)} \left|\frac{\partial(g(x, u_1))}{\partial(x, u_1)}\right|$$

and since in many practical cases U_1 is drawn independently of X and the data

$$\alpha(x, y) = \frac{p(g(x, u_1) \mid \text{data})}{p(x \mid \text{data})} \cdot \frac{j(y)}{q(u_1)j(x)} \cdot \left|\frac{\partial(g(x, u_1))}{\partial(x, u_1)}\right|, \tag{4.10}$$

which is summarized by the shorthand expression:

$$\alpha(x, y) = (\text{posterior ratio}) \times (\text{proposal ratio}) \times (\text{Jacobian})$$

A striking difference of the acceptance probabilities α with the standard case is the Jacobian determinant. This is due to the mechanism devised to make the moves and not to the dimension jump itself. In the standard MCMC for example, we draw samples from the 'target space' from a arbitrary conditional distribution $p(x, dy)$, which then are accepted or rejected based on rule (4.6). In the reversible jump algorithm the samples are drawn indirectly; By first drawing

an auxiliary sample u_1 to cover up the dimension mismatch and then by moving across to the target space with some deterministic function g of u_1 and x. The transition carried out by g concerns spaces of the same dimension, and this is exactly the point where the Jacobian determinant is involved. Its existence is justified because the sample y from the target space is a transformation $g(x, u_1)$ of some point from another space, the properly 'extended' source space. Mathematically and more transparently, the determinant is direct implication of differential calculus' chain rule as shown in (4.9).

4.4 Model Description

Throughout this chapter we use the following introductory notation:

- c denotes a piecewise constant density defined on the unit square.
- The vector of size K, $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_K)$, is a vector of the points, ξ_i from $(0, 1)$, for $i = 1, \ldots, K$. We use $\boldsymbol{\xi}_{-k}$ to denote the vector $(\xi_1, \xi_2, \ldots, \xi_{k-1}, \xi_{k+1}, \ldots, \xi_K)$ which results from $\boldsymbol{\xi}$ after the exclusion of the k-th component.
- The values that c attains in the tiles of constant height will be referred to as levels. We denote the vector of the log-levels by: $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots \eta_K)$.
- $(\boldsymbol{u}, \boldsymbol{v}) = \big((u_1, v_1), (u_2, v_2) \ldots (u_n, v_n)\big)$ denotes a sample from some copula C.

4.4.1 Prior.

A piecewise constant function c can be determined by the number of the consisting pieces, their location on the domain and their levels. A convenient way to define the individual domains of the components (or pieces) of c, is by the use of Voronoi cells. A Voronoi tessellation, generated by the point pattern $\boldsymbol{\xi}$ on a bounded set E, is the partition $E = \cup_{k=1}^{K} E_k(\boldsymbol{\xi})$ into non-overlapping tiles (or cells)

$$E_k(\boldsymbol{\xi}) = \{s \in E : ||s - \xi_k|| \leq ||s - \xi_j|| \text{ for all } j\}$$

where $||\cdot||$ denotes the Euclidean distance. The point ξ_k that generates the corresponding E_k is referred to as its generating point. The purpose of this

partitioning is to create tiles E_k which include all those points s in the domain which are closer to generating point ξ_k than to any other ξ_j. On the plane for the example, the Voronoi partitions are convex polygons. Tiles that share a common edge or even a vertex are deemed to be neighbours. In such cases the common edge is equidistant to the corresponding generating points and perpendicular to the line joining these two points. On the real line, the Voronoi tessellation consists of intervals that have the common endpoint lying in the middle between the two corresponding generating points. In such cases, each tile has two neighbours, apart from those associated with the largest or smallest component of $\{\xi_1, \xi_2, \ldots, \xi_K\}$ which have only one neighbour. Okabe et al. (2000) give a thorough review of concepts, applications and algorithms related to Voronoi diagrams. For a comprehensive survey the reader is also referred to the article by Aurenhammer (1991).

Every vector $\boldsymbol{\xi}$ imposes a neighbouring relationship between the tiles generated by $\xi_i, i = 1, \ldots, K$. We use the notation $\xi_i \underset{\boldsymbol{\xi}}{\sim} \xi_j$ to say that the tiles generated by ξ_i and ξ_j are neighbours. Because the symmetric operation $\underset{\boldsymbol{\xi}}{\sim}$ is $\boldsymbol{\xi}$-dependent the notation should include the vector of points. For brevity however and when the reference to $\boldsymbol{\xi}$ is not vague we will omit it. We may also refer to the indices directly rather to the points themselves, and write $i \sim j$ as an equivalent to $\xi_i \sim \xi_j$. The set of neighbours of the k-th tile is denoted by $\partial_k(\boldsymbol{\xi})$. Hence, $\partial_k(\boldsymbol{\xi}) = \{j : j \sim k\}$. Finally, the neighbourhood system is denoted by $\mathcal{N}(\boldsymbol{\xi}) = \{\partial_k(\boldsymbol{\xi}), k = 1 \ldots K\}$.

When a tile is added to the existing pattern, only the neighbours of the new tile are affected, a property that makes Voronoi tessellations computationally very convenient. The topology of the new pattern (i.e., vertices, tile areas, generating points) changes only in the area consisting from the union of the new tile and its neighbours. This is graphically demonstrated in figure 4.1 which shows an initial set of points in blue color that is updated with the addition of a new point at (0.5, 0.5). It is clear that each of the neighbouring tiles is reduced in size to compensate for the new tile (drawn in red).

Such local updating schemes are central to the implementation of the model that will be presented in the following sections, for reasons of both mathematical and computational convenience, which makes clear the benefits of adopting the Voronoi partitioning approach.

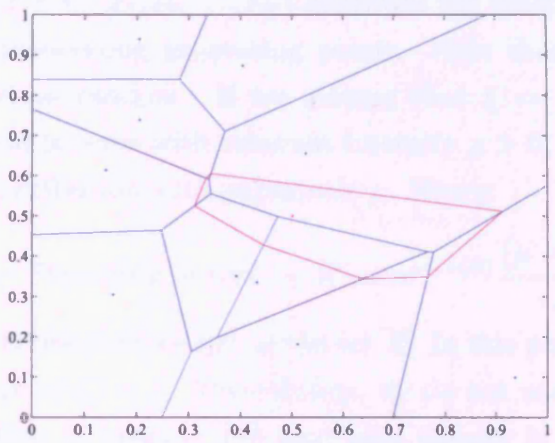


Figure 4.1: Local updating of Voronoi partitions. The addition of the new tile (red colour) leaves its non-neighbouring tiles unchanged.

Suppose now that we have a piecewise constant function,

$$c(s) = \sum_{k=1}^{K} \lambda_k 1_{s \in E_k(\boldsymbol{\xi})}.$$

To determine c, three components are necessary:

- the number K of the components of c
- their position on the domain
- their corresponding levels.

Therefore, the density of our model is equivalent to the joint density of the three random elements mentioned above

$$p(\text{Model}) = p(c) = p(\text{levels, tiles' position, number of tiles})$$

and by the chain rule

$$\begin{aligned} p(c) = {} & p(\text{levels} \,|\, \text{tiles' position, number of tiles}) \\ & \times p(\text{tiles' position} \,|\, \text{number of tiles}) \\ & \times p(\text{number of tiles}). \end{aligned}$$

The random vector $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_K)$ describes the locations of the tiles by specifying the corresponding generating points. Note that the number K of the tiles is treated as random. If we assume that $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_K)$ is a homogenous Poisson process with constant intensity $\mu > 0$, then the counts K, follow a Poisson distribution with parameter μ. Hence:

$$p(K) = \mathbb{P}(\text{number of tiles } = K) = e^{-\mu\,\nu(E)} \frac{\big(\mu\ \nu(E)\big)^K}{K!}, \tag{4.11}$$

where $\nu(E)$ is the Lebesque measure of the set E. In this particular case, where E is the unit square $\nu(E) = 1$. Nevertheless, we do not make the substitution in order to show that in general, the final joint density is independent of the term $\big(\nu(E)\big)^K$.

Conditionally on K, the locations $(\xi_1, \xi_2, \ldots, \xi_K)$ have the same distribution as the order statistics corresponding to K independent random variables uniformly distributed on the domain, the unit square.

$$p(\xi_i, \xi_2, \ldots, \xi_K \mid K) = \frac{K!}{(\nu(E))^K}. \tag{4.12}$$

The levels $\boldsymbol{\lambda} = (\lambda_1, \lambda_1, \ldots \lambda_K)$ are modelled in terms of the distribution of the log-levels $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_K)$ and we assume a Markov Random Field (MRF) to define the joint distribution of $\boldsymbol{\eta}$. MRF theory provides a convenient and consistent way to model context-dependent entities and has been successfully introduced in many fundamental issues of image analysis and computer vision where the pixels have correlated features, such as image restoration Besag (1986), Geman and Geman (1984), edge detection Geman et al. (1990), or surface reconstruction Chou and Brown (1990), Marroquin et al. (1987). This is achieved by characterizing mutual influences among such entities using conditional MRF distributions. In a MRF, the sites (or generating points in our context) in E are related to one another via a neighbourhood system $\mathcal{N}(\boldsymbol{\xi}) = \{\partial_k(\boldsymbol{\xi}), k = 1 \ldots K\}$.

In our problem, the distribution of log-levels is modeled as a MRF with the associated neighborhood system $\mathcal{N}(\boldsymbol{\xi})$. The joint distribution of a MRF is uniquely defined by specifying the conditional distribution (local characteristics) of η_k given its neighbors, $\{\eta_j, j \sim k\}$. In this context we assume a Normal model for the full conditionals of the log-levels $\eta_k, i = 1 \ldots K$:

$$p(\eta_k \mid \boldsymbol{\eta}_{-k}, \tau) \sim N\left(\frac{\sum_{j\sim k} \eta_k}{N_k}, \frac{1}{\tau N_k}\right) \tag{4.13}$$

where $\tau > 0$ is a known hyperparameter and N_k the number of neighbours of point ξ_k.

These specifications for the univariate margins imply a pairwise difference prior for the joint distribution of $\boldsymbol{\eta}$:

$$p(\boldsymbol{\eta} \mid \boldsymbol{\xi}, \tau) \propto \left(\frac{\tau}{2\pi}\right)^{\frac{K}{2}} |W|^{\frac{1}{2}} \exp\Big(-\frac{\tau}{2} \sum_{j \sim k} (\eta_k - \eta_j)^2 \Big), \tag{4.14}$$

where W is a $K \times K$ matrix with elements:

$$w_{kj} = \begin{cases} -1 & \text{if } j \sim k, \\ N_k & \text{if } k = j, \\ 0 & \text{otherwise,} \end{cases}$$

Heikkinen (1998) and Besag et al. (1995). As the authors point out this distribution is improper, but under the presence of some informative data any such impropriety is removed from the posterior. We demand, however, that the full conditionals are proper densities. One of the most important features of the prior in (4.14) is that it remains unchanged when a multiplicative constant is applied on the density c. Functions c and $K_c \times c$ have exactly the same prior distribution for K_c some constant. Therefore, c can be scaled at will, a property which is very useful when there are constraints that must be satisfied, such as integration to one in this particular case.

The conditional distributions (4.11),(4.12) and (4.14) complete the specification of the prior distribution of the piecewise constant function c. The next section discusses the posterior.

4.4.2 Posterior

Suppose now that we observe a random sample $(\boldsymbol{u}, \boldsymbol{v}) = \big((u_1, v_1), (u_2, v_2), \ldots, (u_n, v_n)\big)$, from an unknown copula C and we assume some tiles, chosen at will, in order to specify the arbitrary piecewise constant density c that describes the data. A typical example is illustrated in figure 4.2 where we plot 2000 observations from the copula of the form

$$C(u, v) = \Big[1 + \big((u^{-\alpha} - 1)^{\beta} + (v^{-\alpha} - 1)^{\beta}\big)^{1/\beta}\Big]^{-1/\alpha},$$

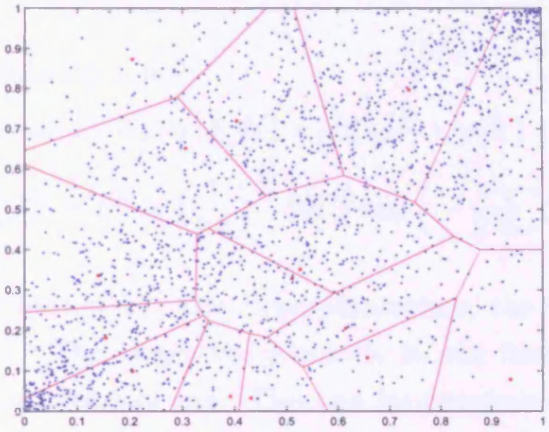


Figure 4.2: A set of data. The domain is partitioned into 15 individual tiles.

with parameters $\alpha = 0.5$, $\beta = 1.5$ overimposed by a set of 15 Voronoi tiles E_k generated by the points ξ_k (plotted in red), $k = 1 \dots 15$.

Because the density c is a piecewise constant function, for any point $s \in E$,

$$c(s) = \sum_{k=1}^{K} \lambda_k 1_{s \in E_k(\boldsymbol{\xi})},$$

and the loglikelihood is expressed as:

$$\begin{aligned} l(\text{data} \mid c) &= \sum_{i=1}^{n} \log(c(u_i, v_i)) \\ &= \sum_{k=1}^{K} \omega_k \log(\lambda_k) \\ &= \sum_{k=1}^{K} \omega_k \, \eta_k \end{aligned} \tag{4.15}$$

where ω_k is the number of datapoints contained in tile E_k,

$$\omega_k = \#\{(u_i, v_i) : (u_i, v_i) \in E_k\}$$

and therefore, $\sum_{k=1}^{K} \omega_k = n$. Our aim now is to sample from the posterior

distribution:

$$\begin{aligned} p(c \,|\, \text{data}) &\propto p(\text{data} \,|\, c) \times p(c) \\ &= p(\text{data} \,|\, c) \times p(\boldsymbol{\eta} \,|\, \boldsymbol{\xi}, K) \times p(\boldsymbol{\xi} \,|\, K) \times p(K) \\ &\propto \prod_{k=1}^{K} \lambda_k^{\omega_k} \times \mu^K \times \left(\frac{\tau}{2\pi}\right)^{\frac{K}{2}} |W|^{\frac{1}{2}} \exp\Big(-\frac{\tau}{2}\sum_{j\sim k}(\eta_k - \eta_j)^2\Big) \end{aligned} \tag{4.16}$$

This joint distribution has two hyperparameters, the Poisson intensity μ and the precision τ. The larger the precision is, the finer the details of the realizations from the posterior are. This can be counterbalanced by setting a small value for μ that controls the grid resolution. A very large resolution is not optimal because it will slow down the convergence of the sampler (described in the next section), in the sense that the scanning of the space is done in very small steps.

To solve the problem of obtaining an approximate sample from the posterior distribution of the form (4.16), we employ a Monte Carlo method which is based on the simulation of a Markov Chain that converges to the desirable distribution. Details of the sampler are discussed in the following section.

4.5 Sampling from the posterior.

One of the parameters in the formulation (4.16) is K, which determines the dimension of c. To allow the Markov Chain to visit spaces of different dimensionality we adopt the reversible jump Markov Chain Monte Carlo framework proposed by Green (1995). It is worth noting that although this type of trans-dimensional problem provides the most common use of RJMCMC, the algorithm itself is more general. Indeed it can be shown to encompass many MCMC type algorithms including standard MCMC.

We choose to incorporate into the algorithm three moves; a dimension preserving move and a pair of birth-death moves that form the reversible jump. As the name suggests the latter are designed to enable the sampler to scan spaces of different dimensionality whereas the dimension preserving move concerns iterations of the Markov chain within the same real space. For the birth or death move, we propose either the addition of an extra tile (and hence an extra component in the piecewise function c) or the deletion of one respectively. In the

remainder of this chapter we use c' to denote the proposed piecewise constant function to be visited by the chain while the current function is c. The three move types are:

1. *Change of a log-level η_k.* In this move we change the value of the density c over a specific tile. An index k is sampled from the discrete uniform distribution on $\{1, \dots K\}$ and the new proposed log-level η'_k is drawn from the uniform distribution on $[\eta_k - \delta, \eta_k + \delta]$, where $\delta > 0$ is a sampler hyperparameter. Since the number and the locations of the consisting components in c do not change, $K' = K$ and $\boldsymbol{\xi}' = \boldsymbol{\xi}$. The log-levels in the neighbourhood of ξ_k change according to a rule described in the next section, whereas $\boldsymbol{\eta}'_{-\partial_k} = \boldsymbol{\eta}_{-\partial_k}$ up to a scale.

2. *Birth of an extra tile.* The move is part of the reversible jump pair and we propose the addition of a new component to the piecewise constant function c. We draw a point from the unit square and it is appended to the existing set of points $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_K)$. The extra tile generated by the new point alters the current Voronoi partition. The dimension of $\boldsymbol{\xi}$ is changed by one and let us denote the new point by $\xi'_{K'} = \xi'_{K+1}$. The log-level associated with $\xi'_{K'}$ is specified in a manner which is outlined later in this section, where we also discuss how the neighbouring log-levels $\boldsymbol{\eta}_{\partial_{K'}(\boldsymbol{\xi}')}$ change in order keep c' to a density.

3. *Death of an existing tile.* That is the reverse of the birth move and suggests that the number of pieces in c are reduced by one. An index k is sampled from the discrete uniform distribution on $\{1, \dots K\}$ and the k-th piece is deleted by omitting point ξ_k from $\boldsymbol{\xi}$ (along with tile E_k) and its corresponding log-value η_k. Hence $\boldsymbol{\xi}' = \boldsymbol{\xi}_{-k}$. To satisfy the density constraint on c', we require that the neighbouring log-levels $\boldsymbol{\eta}_{\partial_k(\boldsymbol{\xi})}$ are properly updated.

In each iteration of the Markov Chain one of the moves is chosen with probability:

$$b_k = \begin{cases} \gamma & \text{if K} \leq \mu \text{ - } 1 \\ \gamma \frac{\mu}{K+1} & \text{otherwise} \end{cases}$$

$$d_k = \begin{cases} 0 & \text{if } K = 2 \\ \gamma \frac{K}{\mu} & \text{if } 2 < K \leq \mu \\ \gamma & \text{if } K > \mu \end{cases}$$

and

$$h_k = 1 - (b_k + d_k)$$

where γ is a constant from $(0, \frac{1}{2})$ that controls the rate each move is proposed with and how often the number of the pieces in c changes. It is easy to see that those probabilities are derived from the likelihood ratios, $b_k = \frac{p(K+1)}{p(K)}$ and $d_K = \frac{p(K)}{p(K+1)}$. Therefore the intuition behind the specification of b_k and d_K is to encourage the chain to scan the space where the counts of the generating points have the greater Poisson likelihood. An extra constraint has been applied on d_k which ensures that when $\boldsymbol{\xi}$ consists of only two points (ξ_1, ξ_2), the probability of deletion is zero.

4.5.1 The dimension preserving move.

In type 1 move, the proposed log-level is drawn as:

$$\eta'_k \sim U(\eta_k - \delta, \eta_k + \delta) \tag{4.17}$$

and we update the neighbouring log-levels $\boldsymbol{\eta}_{\partial_k}$ according to the rule:

$$\eta'_j = \eta_j - \frac{\nu(E_k)}{\nu(E_j)} \cdot \frac{\eta'_k - \eta_k}{N_k}, \quad \text{for } j \in \partial_k. \tag{4.18}$$

This aims to keep the integral of the function $\eta'(s) = \sum_{i=1}^{K} \eta'_i 1_{s \in E_i(\boldsymbol{\xi})}$ for $s \in E$, invariant to the change. Indeed,

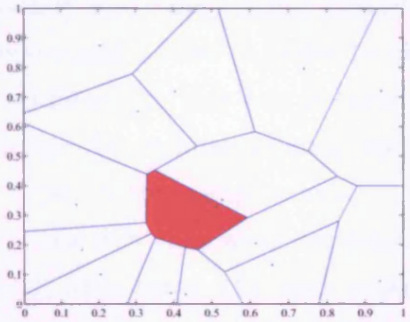
$$\sum_{i=1}^{K} \eta_i \ \nu(E_i) = \sum_{i=1}^{K} \eta'_i \ \nu(E_i).$$

However, the integral of $c'(s) = \sum_{i=1}^{K} \lambda'_i 1_{s \in E_i(\boldsymbol{\xi})}$ will not in general equal to one,

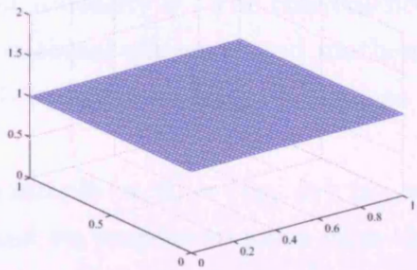
$$\sum_{i=1}^{K} \lambda'_i \ \nu(E_i) \neq 1.$$

To force the move to give proper densities we rescale c' by dividing all levels with the term $\sum_{i=1}^{K} \lambda'_i \ \nu(E_i)$. We are eligible to do so because of the form of the prior (4.14) which is invariant to scale transformation of the levels.

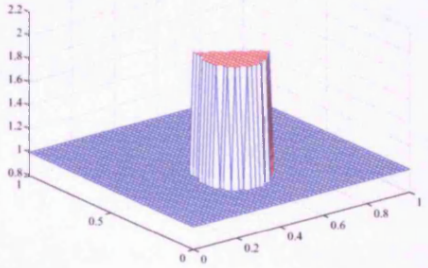
A move 1 type transition from c to c' is described with a simplified but illustrative example in figure 4.3. Suppose we have a vector of 15 generating points $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{15})$ that partitions the domain $[0,1] \times [0,1]$ with the Voronoi tiles as shown in figure 4.3(a). Let us assume that the initial stage of the algorithm has picked up point $\xi_6 = (0.510, 0.329)$ which generates the tile patched with red in the plot. Before any change is proposed the density c is arbitrary taken as jointly uniform, therefore its value is constant and equal to one everywhere on the domain, $c(s) = 1$, for all s. A plot of the density c is shown in figure



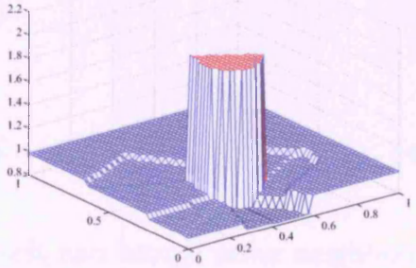
(a) A partitioning of the unit square with some Voronoi tiles.



(b) The initial density c.



(c) The function which results from c after changing a specific tile's level only.



(d) The proposed density c' (after rescaling).

Figure 4.3: A step-by-step illustration of a move 1 type proposal.

4.3(b). It follows that the log-levels are zero, $\eta_k = 0$ for all $k = 1, \ldots 15$. Suppose now that the proposal for η_6 concerns its change from zero which is its current value to $\eta_6' = 0.714$. This proposed log-level corresponds to the level $\lambda_6' = e^{0.714} = 2.042$. Figure 4.3(c) depicts the function which results from the

density c after the level λ_6 has been set to the proposed value λ_6'. It is clear that this function does not integrate to one since there is no compensation for the extra volume gained by the proposal. We choose to change the neighbouring log-levels by employing rule (4.18) aiming in improving the mixing and speeding up the convergence of the chain. Then we rescale all levels so that the function integrates to one. For instance, the level λ_6 after the updating and the scaling is finally set to $\lambda_6' = 2.012$. Note that neighbouring tiles with the smallest areas have undergone the largest deviations from their corresponding initial levels. It is also clear that levels of non-adjacent tiles remain intact up to a scalar. The locally updating scheme described here is one the many (local or non-local) someone could devise in order to construct a density c'. The convenience with local updates like this one is their computational efficiency and mathematical simplicity. As we have already mentioned we intent to keep all changes for all move types in the sampler local.

Suppose now that we observe a random sample $(\boldsymbol{u}, \boldsymbol{v}) = \big((u_1, v_1), (u_2, v_2), \ldots, (u_n, v_n)\big)$, from some unknown copula C and we propose to move from the current c to some new density c', where the latter is the result of a move of type 1 as just described. The distribution in (4.16) implies that the posterior ratio

$$\frac{p(c' \,|\, \text{data})}{p(c \,|\, \text{data})},$$

takes the from:

$$\prod_i \Big(\frac{\lambda_i'}{\lambda_i}\Big)^{\omega_i} \times \exp\Big(-\frac{\tau}{2}\sum_{j,l\in\mathcal{K}}\big[(\eta_l' - \eta_j')^2 - (\eta_l - \eta_j)^2\big]\Big) \tag{4.19}$$

where $\mathcal{K}$ is the set of the indices of all the first and second order neighbouring tiles of tile E_k,

$$\mathcal{K} = \partial_{\{\partial_k(\boldsymbol{\xi})\cup\{k\}\}}(\boldsymbol{\xi}).$$

Finally, the proposal from c to c' is accepted with probability:

$$\alpha(c, c') = \min\left\{1, \frac{p(c' \,|\, \text{data})}{p(c \,|\, \text{data})}\right\}.$$

4.5.2 Reversible jump.

In this section we discuss the elements of the reversible jump pair of moves. We present in detail the birth move only because the other move, death of a tile, is the complete reverse.

For demonstration purposes, assume that the proposed density from the birth has now been accepted. We include a plot of this density in figure 4.4(a). Suppose also that the extra point to be added is $\xi'_{K'} = \xi'_{K+1} = (0.4, 0.3)$ which is shown as a '+' in figure 4.4(b). The new point expands the current vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_K)$ to $\boldsymbol{\xi}' = (\xi_1, \ldots, \xi_K, \xi_{K+1})$. Points $(\xi_1, \ldots, \xi_K)$ are plotted with a dot in figure 4.4(b) where you can also see the corresponding tiles $E_k, k = 1 \ldots K$. When $\xi'_{K'}$ is introduced, tile $E'_{K'}$ is generated causing alterations to the neighbouring tiles $E_k, k \in \partial_{K'}(\boldsymbol{\xi}')$ shown in figure 4.4(b) with dashed lines overimposing $E'_{K'}$. From the same plot is it easy to see the area $\nu(E_k) - \nu(E'_k)$ that each of the tiles $E_k, k \in \partial_{K'}(\boldsymbol{\xi}')$ contributes to $E'_{K'}$.

Once the latter has been included into the partitioning of the domain, we assign a value to represent its associated level $\lambda'_{K'}$. This is done in terms of the proposed log-level $\eta'_{K'}$ which is taken as:

$$\eta'_{K'} = \tilde{\eta}_{K'} + \varepsilon, \tag{4.20}$$

where the perturbation ε is drawn from a distribution with density,

$$f(\varepsilon) = \frac{C_\varepsilon e^{C_\varepsilon \varepsilon}}{(1 + e^{C_\varepsilon \varepsilon})^2}, \tag{4.21}$$

for $C_\varepsilon > 0$ some known hyperparameter, and

$$\tilde{\eta}_{K'} = \sum_{k \in \partial_{K'}(\boldsymbol{\xi}')} \frac{\nu(E_k) - \nu(E'_k)}{\nu(E'_{K'})} \eta_k, \tag{4.22}$$

the weighted sum of the neighbouring log-levels. The weighting in the last relationship expresses the proportional contributions of the neighbouring tiles to the new tile $E'_{K'}$.

When $\eta'_{K'}$ has been specified according to (4.20), the neighbouring log-levels $\eta_k, k \in \partial_{K'}(\boldsymbol{\xi}')$ are updated to the new log-levels according to

$$\eta'_k = \frac{\nu(E_k)}{\nu(E'_k)} \eta_k - \frac{\nu(E_k) - \nu(E'_k)}{\nu(E'_k)} \eta'_{K'}, \tag{4.23}$$

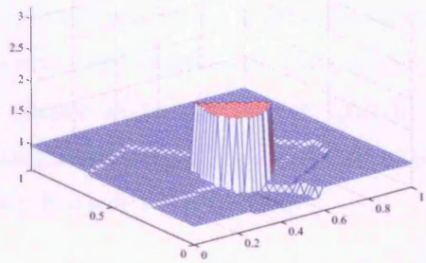
which keeps the integral of the log-densities, before and after the change, the same. Straightforward algebra verifies:

$$\sum_{i=1}^{K} \eta_i \ \nu(E_i) = \sum_{i=1}^{K'} \eta'_i \ \nu(E'_i).$$

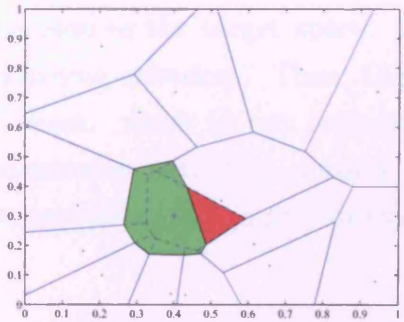
In complete analogy with a move 1 type proposal, the integral of the proposed levels over the domain,

$$\sum_{i=1}^{K'} \lambda'_i \ \nu(E'_i), \tag{4.24}$$

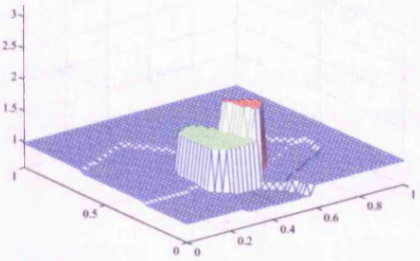
will not always equal to one. Hence, to make this function a proper density we divide it with the term (4.24).



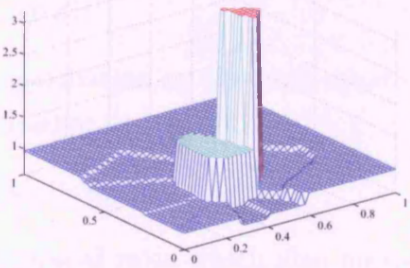
(a) The current density c.



(b) Addition of an extra tile.



(c) The proposed level for the new tile.



(d) The proposed density c'.

Figure 4.4: A step-by-step illustration of a birth move.

The procedure just described is graphically illustrated in figure 4.4. As discussed already, figure 4.4(b) shows the effect of the addition of the new tile

on its neighbours. The next step is to calculate the log-level $\eta_{K'}$ by employing rule (4.20); the achieved level $\lambda'_{K'} = e^{\eta'_{K'}}$ can be seen in figure 4.4(c). Up to this stage every other level has not been changed. This is the task of formula (4.23) which, in combination with the rescaling, ensures that the result of the addition of an extra piece to the piecewise constant function c will always integrate to one. One can see in figure 4.4 how the neighbouring levels change in order to keep the integral invariant.

Before the new point $\xi_{K'}$ is generated, density c is defined by a vector $(\boldsymbol{\lambda}, \boldsymbol{\xi})$ from the real space of dimension $K + K$, $\mathbb{R}^{2K}$. For instance, in the example illustrated in figure 4.4, this vector belongs to $\mathbb{R}^{30}$ (also referred to as source space). On the other side, the proposed density c' belongs to the target space $\mathbb{R}^{32}$. In order to be able to realize the transition from the former to the latter we have to 'complete' the source space with an extra element in such a manner that the extended space has the same dimension as the target space. This is commonly referred to as the 'dimension matching criterion'. Then, the move will concern real spaces of the same dimension. Back in our example, the obvious choice is to employ a 'patch' of dimension two; this consists of the extra generating point $\xi'_{K'}$ and the random perturbation ε jointly drawn from some distribution with density

$$p(\xi'_{K'}, \varepsilon) = p(\xi'_{K'} \mid \varepsilon)\, p(\varepsilon),$$

and when $\xi'_{K'}$, ε are independent (which is the case here):

$$p(\xi'_{K'}, \varepsilon) = p(\xi'_{K'})\, p(\varepsilon).$$

In our context, where $p(\xi'_{K'})$ is the uniform distribution on the unit square and $p(\varepsilon)$ is given by (4.21), the above joint distribution is:

$$p(\xi'_{K'}, \varepsilon) = p(\varepsilon) = f(\eta'_{K'} - \tilde{\eta}_{K'}),$$

and forms part of what Green (1995) calls proposal ratio which also needs the ratio of the probabilities of the cross space moves to be completed. Specifically, the probability to create a new point is given by b_K. The reverse, i.e., the deletion of that point, has probability $\frac{d_{K+1}}{K+1}$. Hence the proposal ratio for birth part of the reversible jump and under the specific sampler specifications is:

$$\frac{d_{K+1}}{(K+1)\, b_K\, f(\eta'_{K'} - \tilde{\eta}_{K'})}. \tag{4.25}$$

When a random $(\xi'_{K'}, \varepsilon)$ has been drawn, vector $(\boldsymbol{\eta}, \boldsymbol{\xi}, \xi, \varepsilon)$ is mapped via a bisection g to a point $(\boldsymbol{\eta}', \boldsymbol{\xi}')$ on the target space. Function g is of deterministic nature and is defined by the modeller with respect to the needs of the problem at hand. In our particular case, g maps all generating points onto themselves. The same applies for all log-levels apart from the proposed $\eta'_{K'}$ and its neighbouringing values which are derived according to (4.20) and (4.23) respectively. Clearly, the vector of proposed log-levels $\boldsymbol{\eta}'$ is a function of the current log-levels and ε. Hence, $\boldsymbol{\eta}' = \boldsymbol{\eta}'(\boldsymbol{\eta}, \varepsilon)$ and the determinant of the jacobian matrix of the transformation yields:

$$\left|\frac{\partial g}{\partial(\boldsymbol{\eta}, \boldsymbol{\xi}, \xi, \varepsilon)}\right| = \left|\frac{\partial \boldsymbol{\eta}'}{\partial(\boldsymbol{\eta}, \varepsilon)}\right| = \prod_{i \in \partial_{K'}(\boldsymbol{\xi}')} \frac{\nu(E_i)}{\nu(E'_i)}. \tag{4.26}$$

Finally the posterior ratio is:

$$\frac{p(c' \mid \text{data})}{p(c \mid \text{data})} =$$

$$\mu \left(\frac{2\pi}{\tau} \frac{|W(\boldsymbol{\xi}')|}{|W(\boldsymbol{\xi})|}\right)^{\frac{1}{2}} \exp\left\{-\frac{\tau}{2} \sum_{k \in \partial_{K'}(\boldsymbol{\xi}')} \left[(\eta'_k - \eta'_{K'})^2 - \frac{1}{2} \sum_{i \in \partial_k(\boldsymbol{\xi}) \setminus \partial_k(\boldsymbol{\xi}')} (\eta_k - \eta_i)^2 \right.\right.$$

$$\left.\left. + \frac{1}{2} \sum_{i \in \partial_k(\boldsymbol{\xi}') \cap \partial_{K'}(\boldsymbol{\xi}')} (\eta'_k - \eta'_i)^2 - (\eta_k - \eta_i)^2 + \sum_{i \in \partial_k(\boldsymbol{\xi}) \setminus \partial_{K'}(\boldsymbol{\xi}')} (\eta'_k - \eta_i)^2 - (\eta_k - \eta_i)^2 \right] \right\}$$

$$\frac{p(\text{data} \mid c')}{p(\text{data} \mid c)}, \tag{4.27}$$

which involves all up to the second order neighbours of the proposed log-level $\eta'_{K'}$ apart from the evaluation of the ratio $\frac{|W(\boldsymbol{\xi}')|}{|W(\boldsymbol{\xi})|}$. To keep all computations in the prior local we follow Heikkinen (1998) who suggests to calculate the previous ratio by taking into account only the first order neighbours of the new point $\xi'_{K'}$. In this manner an approximation $\hat{W}$ (or $\hat{W}'$) of the matrix $W(\boldsymbol{\xi})$ or $(W(\boldsymbol{\xi}'))$ is obtained in exactly the same way as matrix $W(\boldsymbol{\xi})$ or $(W(\boldsymbol{\xi}'))$ but with the assumption that only the tiles $E_k, k \in \partial_{K'}(\boldsymbol{\xi})$ (or $E'_k, k \in \{\partial_{K'}(\xi), K'\}$) exist on domain and that the neighbouring relationship among them is captured by the system $\hat{\mathcal{N}}(\boldsymbol{\xi}) = \{\partial_k(\boldsymbol{\xi}), k \in \partial_{K'}(\boldsymbol{\xi})\}$ (or $\hat{\mathcal{N}}'(\boldsymbol{\xi}') = \{\partial_k(\boldsymbol{\xi}'), k \in \{\partial_{K'}(\xi), K'\}$).

The proposed density c' is accepted with probability:

$$\alpha(c, c') = \min\{1, \mathcal{R}\},$$

for

$$\mathcal{R} = \text{(posterior ratio)} \times \text{(proposal ratio)} \times \text{(Jacobian)}. \tag{4.28}$$

where the ratios and the jacobian as in (4.27), (4.25) and (4.26) respectively.

The death move is constructed as the exact reverse of the birth move. For simplicity, assume that the current vector of generating points is $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_K, \xi_{K+1})$ and we propose the deletion of ξ_{K+1} to obtain the proposed vector $\boldsymbol{\xi}' = (\xi'_1, \ldots, \xi'_K)$. The neighbouring log-levels are set to the new values:

$$\eta'_k = \frac{\nu(E_k)}{\nu(E'_k)}\eta_k + \frac{\nu(E'_k) - \nu(E_k)}{\nu(E'_k)}\eta'_{K+1} \quad \text{for } k \in \partial_{K+1}(\boldsymbol{\xi}), \tag{4.29}$$

to preserve the integral and the associated vector of the levels $\boldsymbol{\lambda}'$ of the proper density λ', is directly derived (after the appropriate rescaling).

The proposal is accepted with probability:

$$\alpha(c, c') = \min\left\{1, \frac{1}{\mathcal{R}}\right\},$$

where $\mathcal{R}$ as in (4.28) having first properly swapped the notation since the proposal of the birth move is the current density for the death move and the current density in the birth is the proposed one for the death step.

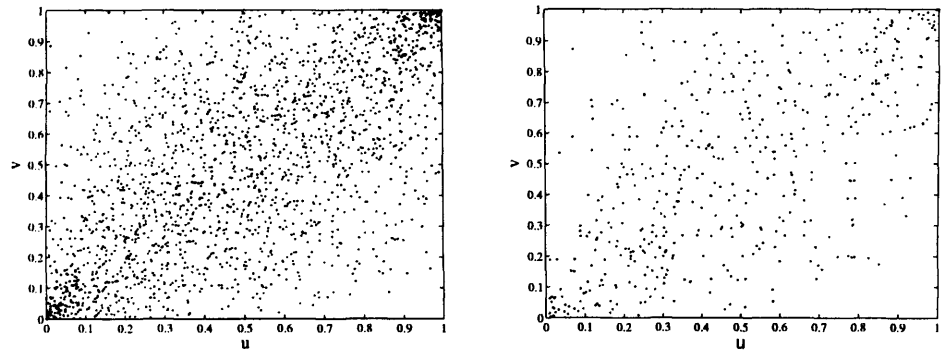
4.6 Applications.

This section demonstrates the performance of the method in three different cases. We sample two datasets of 2000 and 500 observations from a copula of the form (3.13) and we use the approach described in the previous sections to obtain an estimate of the true density. The last example concerns the same dataset as in section (3.3) which consists of 3761 daily readings of the DowJones and FTSE indices. The aim is to estimate the copula density of the joint distribution of the percentage returns of the two indices.

Note that the algorithm simulates a Markov Chain and therefore the samples obtained here are not independent but they are drawn from a distribution, the transition kernel of the Metropolis-Hastings algorithm, which asymptotically converges to the target distribution (in this case, the posterior). In cases where one is interested in the mean, the Markov Chain sample is almost as useful as an

independent one because the mean converges to that of the invariant distribution. It is common however to try eliminate the autocorrelation by introducing a gap between two successive draws. We also let the chain run for a long run (burn-in period) before we start collecting the sample. Although convergence of the transition kernel is achieved only at infinity, in practice the burn-in period serves as a barrier after which the samples can be considered to come from the target distribution. We will refer to this sample as the 'approximate sample'. All these aspects of the sampler (burn-in period, gap) are specified by the practitioner whose tasks also include monitoring of the convergence. One of the main issues of dispute in MCMC was whether valid inference can be based on a single long run or multiple shorter runs are necessary. We adopt the first approach since it is computationally more efficient. Geyer (1992) also, argues that a long enough run suffices.

A scatter plot of the two datasets to test the method is shown in figure 4.5. The generating copula, in both cases, is the member of the bivariate family (3.13) with parameters values $\alpha = 0.5$ and $\beta = 1.5$.



(a) 2000 observations.

(b) 500 observations.

Figure 4.5: The two test datasets.

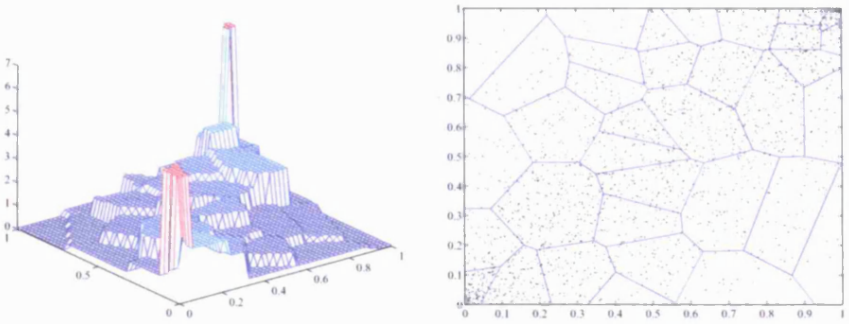
For the first example we use the large dataset of 2000 observations. The simulation of the Markov Chain provide us with an approximate sample from the posterior distribution. The chain is run for an initial 300,000 burn-in iterations before we start collecting one observation every 700 steps until we finally obtain an MCMC sample of size $m = 1000$. This consists of piecewise densities

$\lambda_1, \lambda_2, \ldots, \lambda_m$, each one probably with different numbers of tiles at different positions. The sampler hyperparmeters used here are: $\mu = 40$, $\gamma = 0.4$, $C_\epsilon = 5$, $\delta = 1$ and $\tau = 0.25$ and the initial density where the chain was initialized is the joint uniform which is shown in the upper half of figure 4.3. The latter does not really play a significant role as any other arbitrary chosen initial density also suffices. It is mentioned here only to demonstrate the evolution of the chain.

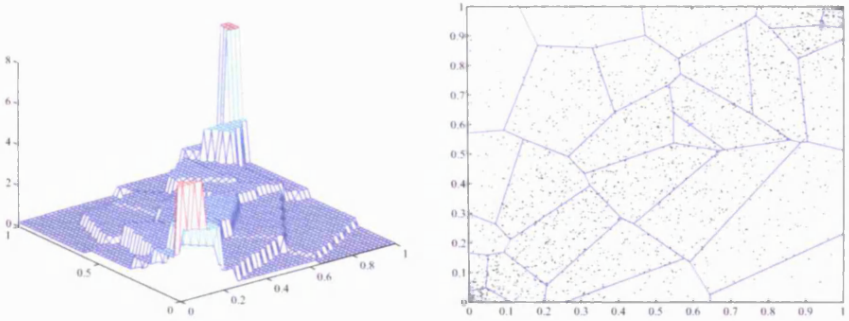
Figure 4.6 includes three typical realizations, c_1, c_{500} and c_{1000}, from the posterior sample. From these examples, it becomes clear that the algorithm behaves as expected, since in areas of high density the number of generating points is larger than in areas of lower density. This is justified by the likelihood which encourages tiles encompassing large number of data points to be assigned high levels and from the prior which acts as a smoother by penalizing large changes between neighbouring tiles. The result is the small tiles appearing in the upper-right and lower-left corners of the Voronoi tessellations included in the right half of figure 4.6. Those tiles have been created in an effort to capture the change in the density under the specifications set by the likelihood and the prior in our model. The large tiles appearing towards the other two corners of the domain have the same justification.

To test the convergence of the transition kernel to the target distribution, we monitor the distribution of the level $c_m(\mathbf{s})$ at six points: $\mathbf{s}_1 = (0.2, 0.2)$, $\mathbf{s}_2 = (0.8, 0.2)$, $\mathbf{s}_3 = (0.8, 0.8)$, $\mathbf{s}_4 = (0.2, 0.8)$, $\mathbf{s}_5 = (0.35, 0.5)$ and $\mathbf{s}_6 = (0.65, 0.5)$. For each one of these points the corresponding level $c_m(\mathbf{s}_i), i = 1 \ldots 6$ is plotted against m, $m = 1 \ldots 1000$ in figure 4.7. Also, in the same figure we draw with red colour the true level that corresponds to those points. The plot shows that we have achieved a fairly good mixing.

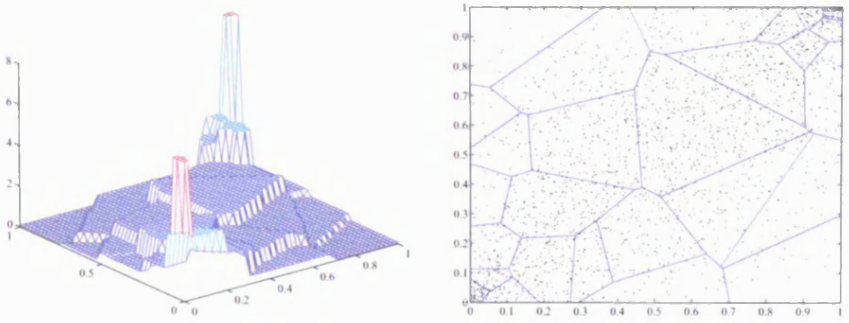
Another feature of the RJMCMC algorithms that we screen, is the acceptance rates for the proposals. The construction of efficient proposals is one of the difficulties encountered when designing reversible jump algorithms and problems of this nature have been reported by many authors; Green and Mira (2001), Brooks et al. (2002), Liu et al. (2001) to name a few. The term 'efficient' refers to proposals which achieve desirable acceptance rates. As noted by Al-Awadhi et al. (2001) inefficient proposals often manifest themselves in very low acceptance rates and typically dimension jumping moves in reversible jump samplers often display much lower acceptance rated than other types of moves



(a) Realization c_1 from the MCMC sample (left) and the corresponding Voronoi tiles (right).

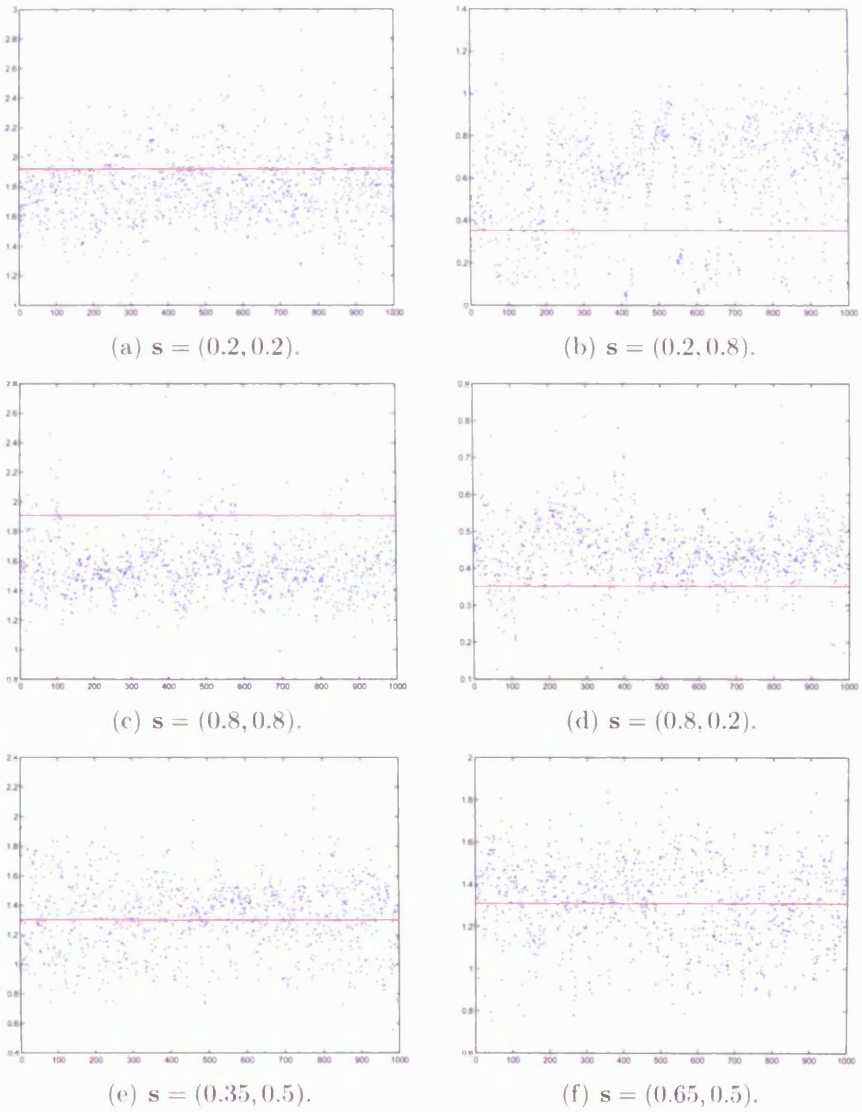


(b) Realization c_{500} from the MCMC sample (left) and the corresponding Voronoi tiles (right).



(c) Realization c_{1000} from the MCMC sample (left) and the corresponding Voronoi tiles (right).

Figure 4.6: Dataset of size $n = 2000$: Typical examples of some realizations from the MCMC sample. Scatter plots of the observations are included in the right half of the figure.



(a) $\mathbf{s} = (0.2, 0.2)$. (b) $\mathbf{s} = (0.2, 0.8)$.

(c) $\mathbf{s} = (0.8, 0.8)$. (d) $\mathbf{s} = (0.8, 0.2)$.

(e) $\mathbf{s} = (0.35, 0.5)$. (f) $\mathbf{s} = (0.65, 0.5)$.

Figure 4.7: Monitoring the convergence of the transition kernel of Markov chain to the stationary distribution. This plot shows the value of $c_m(\mathbf{s})$ for various points $\mathbf{s}$ and while $m = 1 \ldots 1000$. The red continuous line is the true level at those points.

where the dimension is held fixed. Too low acceptance rates essentially lead to poor mixing which in turn leads to slow convergence and increased autocorrelation in the Markov chain. For standard MCMC methods acceptance rates that are too high are also suboptimal, because they are often achieved by proposing states that are close to the existing state. In such cases autocorrelation is also high and thus convergence is slow. Roberts et al. (1997) have found that under certain regularity conditions the optimal acceptance rate for random walk[1] Metropolis algorithms is 0.23 for transitions across the same dimension.

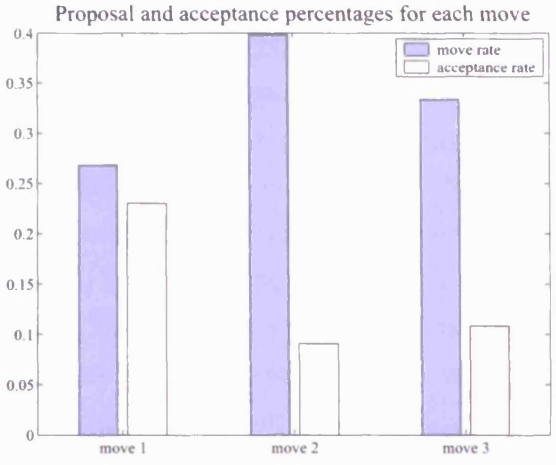


Figure 4.8: Move rates (in blue colour) and corresponding acceptance rates (in brown colour).

Figure 4.8 plots the proportions, in percentages, for each of the three moves across the MCMC iterations after the burn-in period. In the same plot we have included the corresponding acceptance rates (also in percentages).

One of the main features of Bayesian methods is that we obtain a whole sample from the target distribution $p(c \mid \text{data})$ as opposed to frequentistic approaches which give point estimates. Observations from the posterior are illustrated in figure 4.6. A possible candidate for a 'Bayesian point estimate' which we use as an estimator for the unknown density, is the posterior mean

$$\mathbb{E}(c \mid \text{data})$$

[1] Given X, this algorithm proposes a move to $Y = X + \sigma W$, where W is a standard Normal random variable, and σ therefore regulates the size of the proposed jump.

which is estimated by the piecewise posterior sample mean

$$\hat{c}_m(\mathbf{x}) = \frac{\sum_{j=1}^{m} c_j(\mathbf{x})}{m},$$

where $\mathbf{x}$ some evaluation points on the domain.

Figure 4.9 shows a plot of the posterior mean of the MCMC sample for the dataset of 2000 observations, evaluated on a 50×50 grid of points on the unit square. Note that although the individual samples are piecewise constant functions, the mean appears to be a smooth function. It is clear also that the sampler shows satisfactory adaptivity and follows the rapid changes of the density in the left-bottom and top-right corners of the domain. We have to say here that the true density goes asymptotically to infinity as both u and v tend to zero or one, something which is not precisely illustrated in figure 4.9(a).

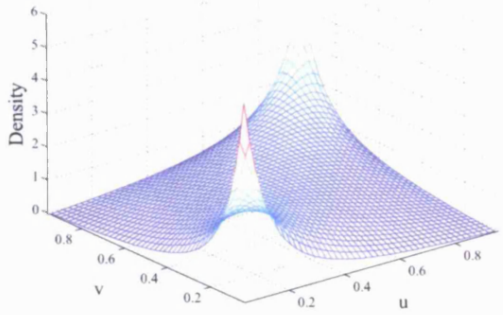
Other interesting conclusions drawn from the MCMC sample concern for example, the marginal distribution $p(c(\mathbf{x}))$ of the density values at some predefined point $\mathbf{x}$. Figure 4.10 illustrates the estimated densities of the levels $c(\mathbf{s}_i)$ at the six control points $\mathbf{s}_i$. The estimates were obtained by constructing a sample of 1000 observations where each observation is the value each posterior density $c_m, m = 1, \ldots, 1000$ attains at the control point $\mathbf{s}_i$. Kernel density estimation on these samples yields the densities illustrated in figure 4.10. The red dot on the horizontal axis gives the value of the true density at the particular points.

Similarly, to estimate the posterior distribution of the number of tiles, we retrieve the dimension of each density c_m and obtain a sample of size 1000. The estimated posterior density of K is plotted in figure 4.11

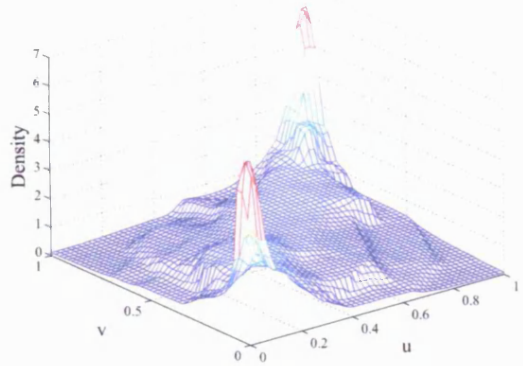
The estimated posterior mean density from the simulation with the small dataset of size $n = 500$ is shown in figure 4.9(c). The hyperparameter values and the evaluation grid are exactly the same as those for $n = 2000$. As expected, restoration of the original density improves as the sample size n increases.

The last example concerns the same data as the ones used in chapter 2. The sample in this case consists of the daily observations of two financial indices, the Dow-Jones Eurostoxx and the FT-all for a period of 3761 days. A scatter plot of the percentage returns is shown in figure 4.12 where also included is a scatter plot of the ranked percentage returns.

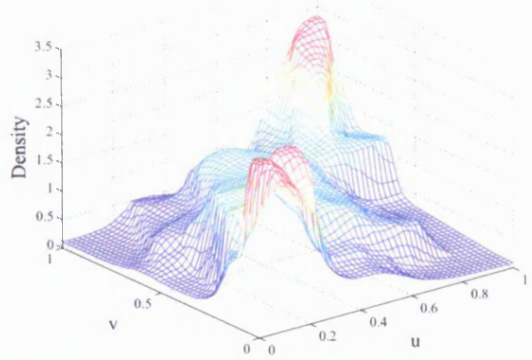
To obtain the ranks of the observations we use empirical margins. We want to estimate the density of the point pattern in figure 4.12(b). The posterior



(a) True density

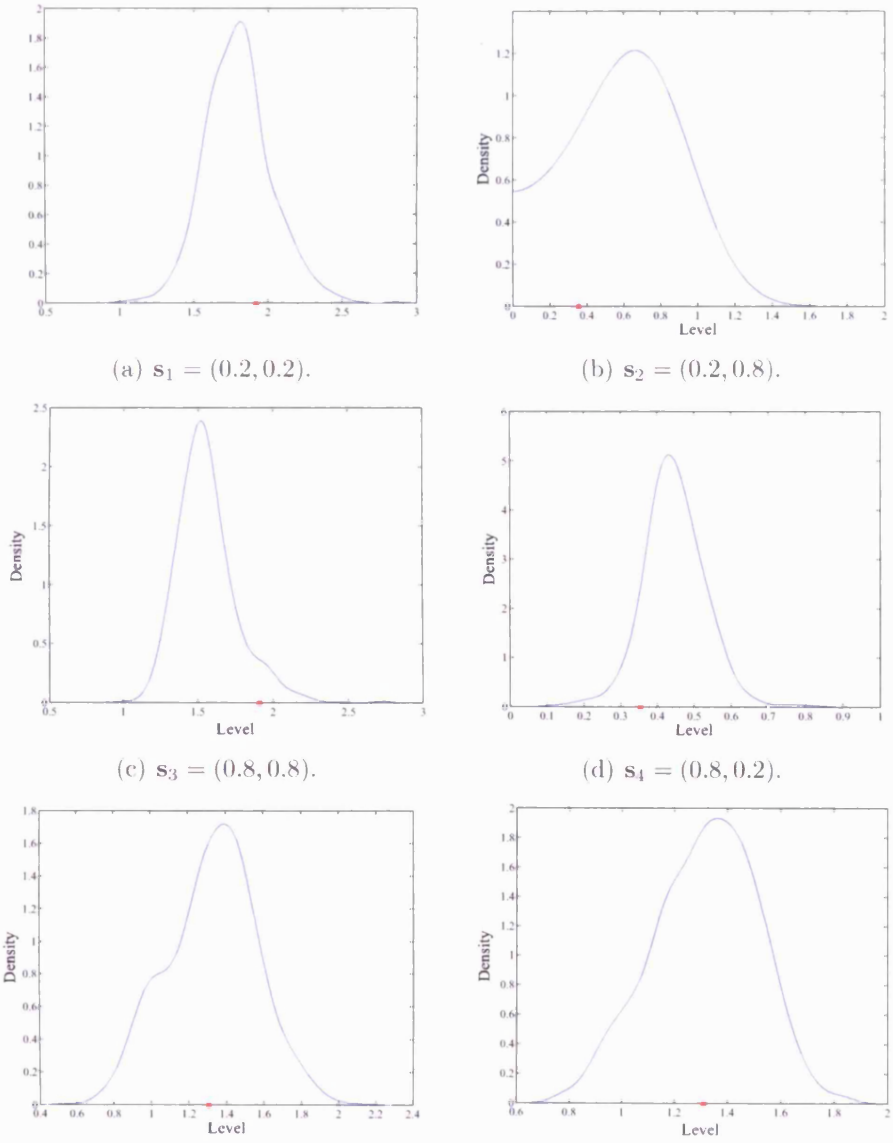


(b) Dataset of size $n = 2000$: Estimated posterior mean density.



(c) Dataset of size $n = 500$: Estimated posterior mean density.

Figure 4.9: Perspective plots of the true density and its estimates for the two datasets.



(a) $\mathbf{s}_1 = (0.2, 0.2)$.

(b) $\mathbf{s}_2 = (0.2, 0.8)$.

(c) $\mathbf{s}_3 = (0.8, 0.8)$.

(d) $\mathbf{s}_4 = (0.8, 0.2)$.

(e) $\mathbf{s}_5 = (0.35, 0.5)$.

(f) $\mathbf{s}_6 = (0.65, 0.5)$.

Figure 4.10: Marginal densities at six control points $\mathbf{s}_i$, $i = 1 \ldots 6$. The red dot gives the true value for the corresponding point.

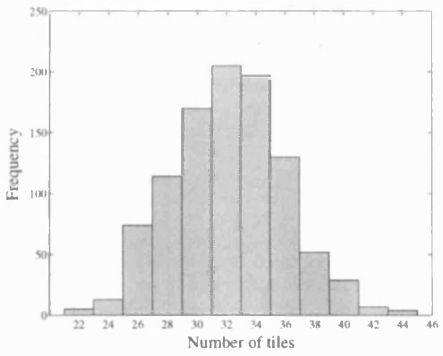
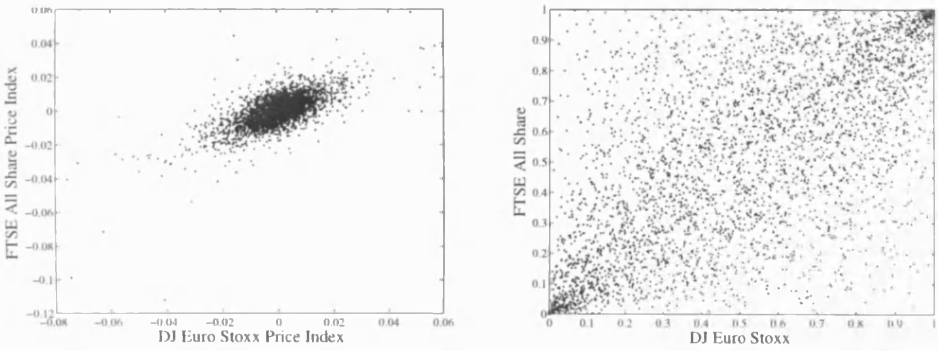


Figure 4.11: Estimated posterior mean density of the number of tiles, K.



(a) Scatter plot of the actual observations.

(b) Scatter plot of the ranks of the observations.

Figure 4.12: Financial indices dataset: Scatter plots of the actual daily percentage returns (left) and the ranked ones (left).

mean of an MCMC sample of size 1000 with $\mu = 50, \gamma = 0.4, C_\varepsilon = 10, \delta = 1$ and $\tau = 1.5$ along with the corresponding contours plot is shown in the upper half of figure 4.13. Pointwise posterior standard deviations are plotted in the lower half of the same figure. As expected, areas of rapid change of the density show larger variability.

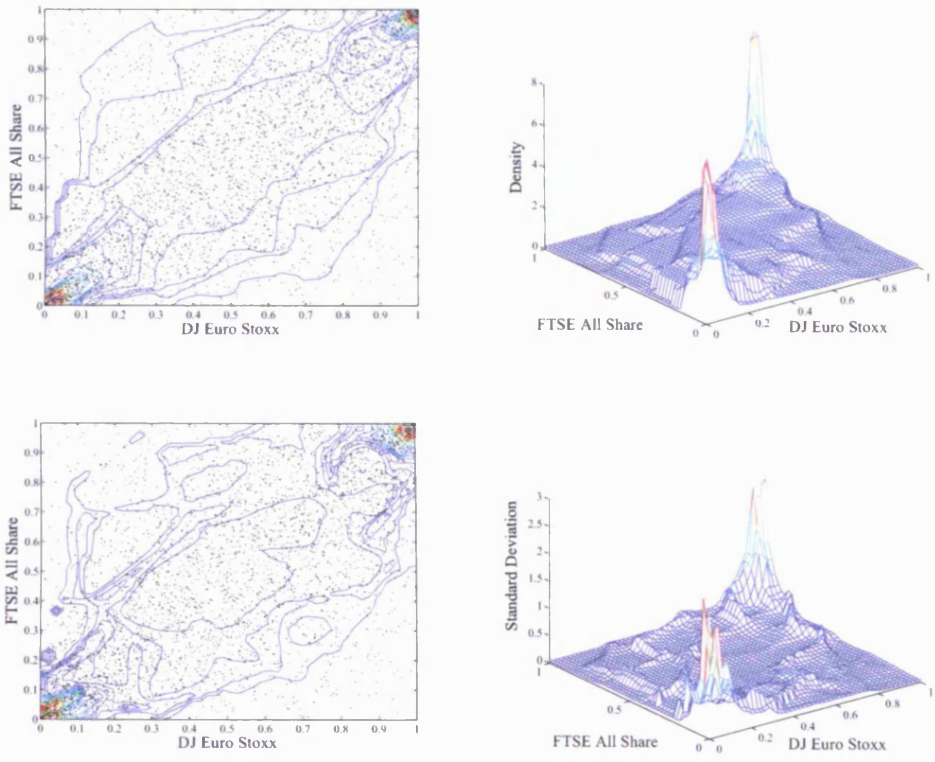


Figure 4.13: Contour and perspective plots of the estimated posterior mean for the sample of the percentage returns of the two indices. The data points are included in the contour plots. Pointwise posterior standard deviations are illustrated in the lower half.

Chapter 5

Non-parametric Bayesian Estimation Under Shape Constraints

5.1 Introduction

This chapter presents a methodology, based on quadratic splines, for the estimation by of the unobservable generator ϕ of an Archimedean copula

$$C(u, v) = \phi^{-1}(\phi(u) + \phi(v)), \quad u, v \in (0, 1).$$

The problem falls in the wider area of shape preserving techniques since ϕ is by definition convex and decreasing. We implement a Bayesian approach for the estimation of the generator ϕ under the assumption that the parametric form is unknown, based on the procedure covered in chapter 4. Specifically, we make the assumption that ϕ'' is a random step function with some prior distribution and we derive the posterior distribution of ϕ'' under the presence of some data drawn from an unknown Archimedean copula. The samples from the posterior are functions with different numbers of steps at different locations, that also attain different values which we call heights or levels. Given ϕ'', we work 'backwards' to recover the two antiderivatives, ϕ' and ϕ. Under the assumption that ϕ'' is a known step function, ϕ and ϕ' are deterministic quadratic and linear splines respectively. Finally, to estimate the theoretical generator and its derivatives we use the corresponding sample mean of the posterior sample.

In order to determine ϕ from ϕ'' we need two constraints. The first is given by the definition, $\phi(1) = 0$, and the other is set by fixing the scale of ϕ. That

is, we choose an arbitrary point x_1 and map it to a positive value chosen at will. This constraint derives from the invariance of Archimedean copulas to scale transformations of the generator.

The posterior sample is obtained by means of a Markov Chain Monte Carlo simulation. The implied joint prior distribution on the step function is the single-dimensional counterpart of the prior used in chapter 4 for the estimation of a bivariate copula density, with some minor changes. The sampler too, has undergone some slight alterations that make it simpler and accommodate the needs of the problem more efficiently.

In section 5.2 we formulate the posterior distribution and we continue in section 5.3 with the description of the MCMC sampler used to obtain an approximate sample from this distribution. The methodology is illustrated using the datasets (simulated or real word) from the previous chapter.

5.2 Formulation of the posterior

We present a Bayesian model for the simultaneous estimation of ϕ and its first two derivatives by assuming that ϕ'', defined on a closed interval $[x_1, x_{K+1}]$, is a step function.

Suppose that ϕ'' has K steps defined on the points $x_1 < x_2 < \cdots < x_{K+1}$. The k-th step is determined by the location x_k and takes the value a_k on the subinterval $[x_k, x_{k+1})$ for $k = 1, 2, \ldots, K$. We construct a prior for the step function by proposing a distribution for the number of steps K, the positions $\boldsymbol{x} = (x_2, x_3, \ldots, x_K)$ given K and the loglevels (or log-heights) $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_K)$, given K and $\mathbf{x} = (x_1, x_2, \ldots, x_K, x_{K+1})$. Hence,

$$\begin{aligned} p(\phi'') &= p(\mathbf{x}, \boldsymbol{\eta}, K) \\ &= p(K) \times p(\boldsymbol{x} \mid K) \times p(\boldsymbol{\eta} \mid \mathbf{x}, K) \end{aligned}$$

Note that the two endpoints x_1, x_{K+1} are considered fixed and known. A typical realisation of $\eta = \log(\phi'')$ is shown in figure (5.1).

In analogy with section 4.4.1, we assume that the $K - 1$ change points $(x_2, x_3, \ldots, x_K)$ is a sequence from a Poisson process with known and constant

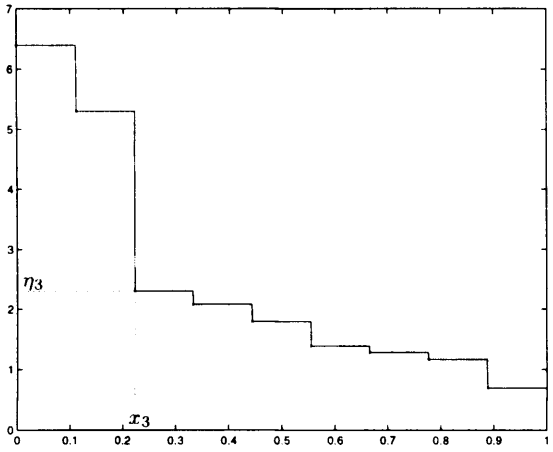


Figure 5.1: A step function.

intensity $\mu > 0$. Therefore, the counts of the steps, K are distributed according to a Poisson distribution

$$p(K) = e^{-L\mu}\frac{(L\mu)^{K-1}}{(K-1)!}.$$

where L is the length $x_{K+1} - x_1$. Conditionally on K, the locations $(x_2, x_3, \ldots, x_K)$, are distributed as the order statistics of $K-1$ points uniformly distributed on $[x_1, x_{K+1}]$,

$$p(x_2, x_3, \ldots, x_K \mid K) = \frac{(K-1)!}{L^{K-1}}$$

The log-levels $\boldsymbol{\eta}$ given $\mathbf{x}$ and K are modelled in terms of a Markov Random Field. Specifically, the full conditionals $p(\eta_k \mid \boldsymbol{\eta}_{-k})$ are taken as a Normal distribution with mean the average log-level of the neighbouring steps,

$$\eta_k \mid \boldsymbol{\eta}_{-k} \sim N\Big(\frac{\sum_{j\sim k}\eta_k}{N_k}, \frac{1}{\tau N_k}\Big), \tag{5.1}$$

where $\tau > 0$ is a hyperparameter and N_k the number of neighbours of the k-th step. Note that two steps defined on subintervals with a common bound are deemed neighbours. Hence, $N_k = 2$ for all $k \neq 1, K$, otherwise $N_k = 1$.

We saw in chapter 4 that the conditionals of the form (5.1) imply the joint distribution:

$$p(\boldsymbol{\eta} \mid \mathbf{x}, \tau) \propto \Big(\frac{\tau}{2\pi}\Big)^{\frac{K}{2}} |W|^{\frac{1}{2}} \exp\Big(-\frac{\tau}{2}\sum_{j\sim k}(\eta_k - \eta_j)^2\Big), \tag{5.2}$$

Suppose now that a dataset $(\boldsymbol{u}, \boldsymbol{v}) = \big((u_1, v_1), (u_2, v_2), \ldots, (u_n, v_n)\big)$ is drawn from some Archimedean copula with generator ϕ. The log-likelihood is given by

$$l(\text{data} \mid \phi) = \log(p(\text{data} \mid \phi)) = \sum_{i=1}^{n}\sum_{j=1}^{n} \log(\phi''(C(u,v))) + \sum_{i=1}^{n} \log(-\phi'(u)) + \sum_{j=1}^{n} \log(-\phi'(v)) - 3\sum_{i=1}^{n}\sum_{j=1}^{n} \log(-\phi'(C(u,v))),$$

and the posterior joint distribution turns out to be:

$$p(\phi'' \mid \text{data}) \propto \mu^{K-1} \left(\frac{\tau}{2\pi}\right)^{\frac{K}{2}} |W|^{\frac{1}{2}} \exp\Big(-\frac{\tau}{2}\sum_{j\sim k}(\eta_k - \eta_j)^2\Big)\, p(\text{data} \mid \phi'') \quad (5.3)$$

The prior of the posterior in (5.3) is identical to the distribution used for the computation of the posterior in (4.16), chapter 4. We now however exploit the extra information that the generic functional form of the Archimedean family is available up to ϕ and its derivatives. The use of a step function as a representation for ϕ'' is a convenient framework for our investigations. Specifically, if

$$\phi''(x) = \sum_{k=1}^{K} a_k 1_{x\in[x_k, x_{k+1})}$$

and we express the generator as the quadratic spline

$$\phi(x) = \sum_{k=1}^{K} \Big\{\frac{a_k}{2}(x - x_k)^2 + \beta_k(x - x_k) + \gamma\Big\} 1_{x\in[x_k, x_{k+1}]}$$

then, for $(\beta_1, \beta_2, \ldots, \beta_K)$ and $\phi'(1)$ all negative, the derivative ϕ' is a negative function and ϕ decreasing. In this way we satisfy one of the generator properties by controlling the value of ϕ at one and the constant coefficients of the linear spline only. Furthermore, to ensure that ϕ'' is positive, and hence ϕ convex, we use a log-transformation of the form $\eta_k = \log(a_k)$.

5.3 Sampling from the posterior

Our inferences are based on an approximate sample from the posterior, drawn by simulating a reversible jump Markov chain that has the posterior distribution

as its invariant distribution. In the spirit of section (4.5) the sampler consists of three moves which suggest a number of possible changes on the step function $n = \log(a)$. These proposals concern three types of moves namely, the change of a step height, the birth of a step or the removal of one. We use the notation η^* to denote a realisation obtained from such changes. Quantities in η, such as the number K of the steps or their heights, that have undergone alterations due to these proposals are denoted in an analogous manner.

If η has K steps and b_K, d_K and h_K are the probabilities of the birth, death, or height change moves respectively, we start by randomly selecting one of the three moves according to:

$$b_K = \gamma \min\left\{1, \tfrac{p(K+1)}{p(K)}\right\}, \quad d_K = \gamma \min\{1, \tfrac{p(K)}{p(K+1)}\}, \quad h_K = 1 - (b_K + d_K)$$

where $\gamma \in (0, \frac{1}{2})$ is a known sampler hyperparameter that controls the mixing.

When a change of a step height is proposed the dimensionality of ϕ is preserved. For this type of move we choose an index k from the discrete uniform distribution on $\{1, \dots K\}$ and then we suggest the replacement of η_k by

$$\eta_k^* \sim U[\eta_k - \delta, \eta_k + \delta],$$

where $\delta > 0$ is a known sampler hyperparameter. The number and the position of the steps remains invariant to the proposal, hence $K^* = K$ and $\mathbf{x}^* = \mathbf{x}$. The same applies for the remaining log-heights, therefore $\boldsymbol{\eta}^*_{-k} = \boldsymbol{\eta}_{-k}$. Suppose for example that $\log(\phi'')$ is the function in blue colour in figure (5.2) and consists of 9 steps. For illustration purposes, assume that the sampler has picked up the index $k = 2$ and the associated step height $\eta_2 = 5.29$ is updated to $\eta_k^* = 5.0$ as shown in the plot. From each of the vectors $\boldsymbol{\eta} = \log(\boldsymbol{a})$ and $\boldsymbol{\eta}^* = \log(\boldsymbol{a}^*)$ we can retrieve the corresponding quadratic spines,

$$\phi(x) = \sum_{k=1}^{K} \left\{ \frac{a_k}{2}(x - x_k)^2 + \beta_k(x - x_k) + \gamma \right\} 1_{x \in [x_k, x_{k+1})}$$

and

$$\phi^*(x) = \sum_{k=1}^{K} \left\{ \frac{a_k^*}{2}(x - x_k)^2 + \beta_k^*(x - x_k) + \gamma^* \right\} 1_{x \in [x_k, x_{k+1})}$$

respectively, under the constraints of continuity for ϕ, ϕ^* and their first derivatives at the change points $(x_2, x_3, x_4, \dots, x_K)$, while $\lim_{x \to 1} \phi(x) = \lim_{x \to 1} \phi^*(x) =$

0 and given that $\phi(x_1), \phi^*(x_1)$ are set to a common fixed value. Specifically for $\phi(x)$, we retrieve the coefficients as the solution of the $2K \times 2K$ linear system,

$$\begin{aligned}
&\frac{a_1}{2}(x_2 - x_1)^2 && + \beta_1(x_2 - x_1) && + \gamma_1 && = \gamma_2,\\
&\frac{a_2}{2}(x_3 - x_2)^2 && + \beta_2(x_3 - x_2) && + \gamma_2 && = \gamma_3,\\
&\vdots && \vdots && && \vdots\\
&\frac{a_{K-1}}{2}(x_K - x_{K-1})^2 && + \beta_{K-1}(x_K - x_{K-1}) && + \gamma_{K-1} && = \gamma_K
\end{aligned}$$

$$\begin{aligned}
&a_1(x_2 - x_1) && + \beta_1 && && = \beta_2,\\
&a_2(x_3 - x_2) && + \beta_2 && && = \beta_3,\\
&\vdots && \vdots && && \vdots\\
&a_{K-1}(x_K - x_{K-1}) && + \beta_{K-1} && && = \beta_K,
\end{aligned}$$

$$\begin{aligned}
&&& && \gamma_1 && = \phi(x_1),\\
&\frac{a_K}{2}(1 - x_K)^2 && + \beta_K(1 - x_K) && + \gamma_K && = 0,
\end{aligned}$$

with respect to $\beta_1, \beta_2, \ldots, \beta_K$ and $\gamma_1, \gamma_2, \ldots, \gamma_K$. Similarly for $\phi^*(x)$. A plot of the generator before and after the change is included in figure (5.2). Because all log-heights apart from that of the k-th step remain the same during the transition, $\eta_k = \eta_k^*$. However the constants and the first order[1] coefficients of the quadratic spline are changed hence, $\beta \neq \beta^*$ and $\gamma \neq \gamma^*$. Given a dataset from an unknown Archimedean copula C we evaluate the likelihood ratio

$$\frac{p(\text{data} \mid \phi^*)}{p(\text{data} \mid \phi)}$$

and the proposed replacement is accepted with probability:

$$\begin{aligned}
\alpha(\phi, \phi^*) &= \min\{1, \text{posterior ratio}\}\\
&= \exp\Big[-\frac{\tau}{2}(\eta_k^* - \eta_k) \sum_{j \sim k} (\eta_k^* + \eta_k - 2\eta_j) \Big] \frac{p(\text{data} \mid \phi^*)}{p(\text{data} \mid \phi)}. \qquad (5.4)
\end{aligned}$$

In a birth type move we propose the inclusion of an extra step in $\log(\phi''(x))$ by selecting its position x^* uniformly on $[x_1, x_{K+1}]$ and let us suppose that

[1]The highest power in a (univariate) polynomial is called its order

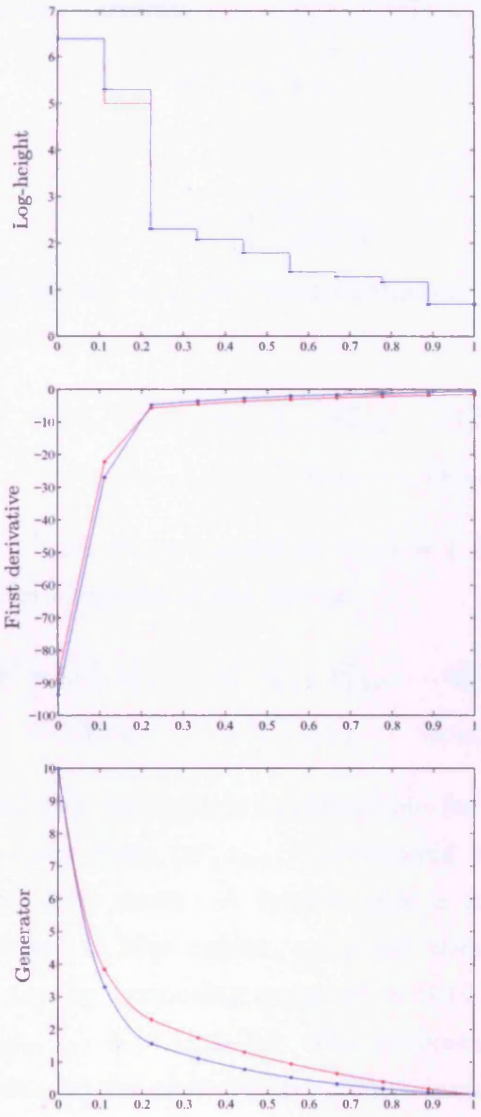


Figure 5.2: Illustration of the height-change move. The top plot shows a step function $\log(\phi''(x))$ in blue color and we propose to replace the height of a step with the new value shown in red. The middle and bottom figures are plots of the corresponding linear and quadratic splines respectively. The dots correspond to the values of the piecewise polynomials that consist the step functions and the splines, evaluated at the bounds of their individual domains.

this falls in an interval $[x_k, x_{k+1})$ where $\log(\phi''(x))$ takes the value η_k. The new step height on the interval $[x^*, x_{k+1})$ is chosen as the current log-height η_k perturbated by a random variable,

$$\eta^* = \eta_k + \varepsilon, \tag{5.5}$$

where ε has density

$$f(\varepsilon) = \frac{C_\varepsilon \mathrm{e}^{C_\varepsilon \varepsilon}}{(1 + \mathrm{e}^{C_\varepsilon \varepsilon})^2},$$

and $C_\varepsilon > 0$ is another known sampler hyperparameter. The proposed function η is defined by a vector

$$\begin{aligned} \boldsymbol{x}^* &= (x_1^*, x_2^*, \ldots, x_k^*, x_{k+1}^*, x_{k+2}^* \ldots, x_{K+2}^*) \\ &= (x_1, x_2, \ldots, x_k, x^*, x_{k+1} \ldots, x_{K+1}) \end{aligned}$$

of $K^* = K + 2$ points, with steps located at x_j^*, $j = 1, 2, \ldots, K+1$ and corresponding log-height the elements of the vector

$$\begin{aligned} \boldsymbol{\eta}^* &= (\eta_1^*, \eta_2^*, \ldots, \eta_k^*, \eta_{k+1}^*, \eta_{k+2}^* \ldots, \eta_{K+2}^*) \\ &= (\eta_1, \eta_2, \ldots, \eta_k, \eta^*, \eta_{k+1} \ldots, \eta_{K+1}). \end{aligned}$$

Note that the proposal does not include modifications for the neighbouring steps and all subintervals apart from $[x^*, x_{k+1})$ correspond to heights that remain unaffected during the birth move. A typical case is included in figure (5.3) where the top plot shows in blue colour, a current step function $\eta = \log(\phi'')$. We generate a new step by proposing point $x^* = 0.16$ which falls within the existing subinterval $[x_2, x_3) = [0.11, 0.22)$. The proposed replacement function η^*, shown in red colour in the same figure, differs from η only on $[0.16, 0.22)$ where it takes the value $\eta^* = 4.0$ drawn according to rule (5.5).

Following the analysis in chapter (4), the proposal ratio for this sampler turns out to be:

$$\frac{d_{K+1}(x_{K+1} - x_1)}{(K+1)\, b_K\, f(\eta^* - \eta_k)} \tag{5.6}$$

and the determinant of the jacobian matrix of the transformation equals to one. The proposed generator is accepted with probability

$$\alpha(\phi^*, \phi) = \{1, R\}$$

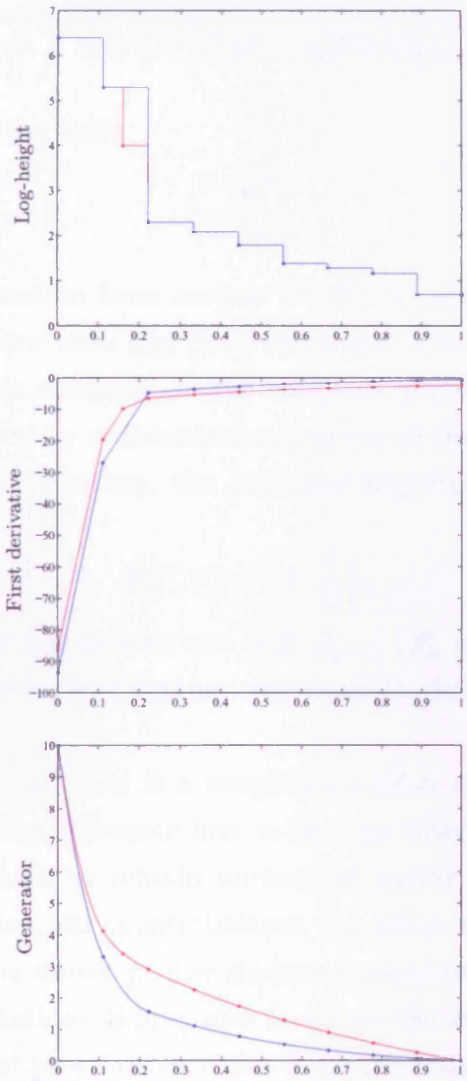


Figure 5.3: Illustration of the birth move. The top plot shows a step function $\log(\phi''(x))$ in blue color which is changed by the addition of the extra step shown in red. The middle and bottom figures are plots of the corresponding linear and quadratic splines respectively. The dots correspond to the values of the piecewise polynomials that consist the step functions and the splines, evaluated at the bounds of their individual domains.

for $R = \{\text{prior ratio} \times \text{likelihood ratio} \times \text{proposal ratio} \times \text{jacobian}\}$ where the prior ratio is:

$$\frac{p(\eta^*)}{p(\eta)} = \mu \left(\frac{2\pi}{\tau} \frac{|W(\boldsymbol{x}^*)|}{|W(\boldsymbol{x})|} \right)^{\frac{1}{2}} \exp \left\{ -\frac{\tau}{2} \left[(\eta^* - \eta_j)^2 + (\eta_{j+1} - \eta^*)^2 - (\eta_{j+1} - \eta_j)^2 \right] \right\}.$$

For the evaluation of the ratio

$$\frac{|W(\boldsymbol{x}^*)|}{|W(\boldsymbol{x})|}, \tag{5.7}$$

we use a the approximation from section (4.5.2) by considering the steps adjacent to the proposed one only and (5.7) turns out to be 3/2.

The death move is designed as the complete reverse. The candidate for deletion step, is selected by uniformly drawing one of the points $(x_2, x_3, \ldots x_K)$. With proper switch in labelling, the proposed step function is accepted with probability

$$\alpha(\phi^*, \phi) = \left\{ 1, \frac{1}{R} \right\}.$$

For all moves, when ϕ^* has at least one of $\beta_1^*, \beta_2^*, \ldots, \beta_K^*$ and $\phi'^*(1)$ non-negative, the proposal is rejected without further consideration since the likelihood $p(\text{data} \mid \phi^*)$ is zero.

The sampler just described is a simplified version of the methodology discussed in chapter 4. The difference here is that the integral of the step function $\log(\phi''(x))$ does not have to remain unchanged during all transitions. In the density estimation case, we counterbalance the effect of the proposals by incorporating within the moves proper updating schemes for the neighbours, in order to keep computations simple and facilitate the mixing. A similar technique could be followed here, however, this is not recommended in this particular case since the results appear to be unsatisfactory. The sampler is encouraged to create more steps in areas of rapid change than in flat ones. When we adopt the updating procedure of chapter 4 and a location $x^* \in [x_k, x_{k+1})$ is proposed for the new step defined on $[x^*, x_{k+1})$ with log-height η^*, the existing value η_k of the subinterval $[x_k, x^*)$ is changed to η_k' to balance out the added volume, $(x_{k+1} - x^*)\eta^*$. If η^* is larger than η_k the adjacent step is set to a value $\eta_k^* < \eta_k$. The shorter the interval $[x_k, x^*)$ the smaller the updated value and the bigger the change are. Figure (5.4) shows a realisation demonstrating the case. The

plotted step function was not penalised in the likelihood and survived in the posterior. Note the striking change on the step heights in the left part of the plot. The updating rule does not imply any changes for the height of the subinterval $[x_{k+1}, x_{k+2})$ because its length remains the same during the proposal.

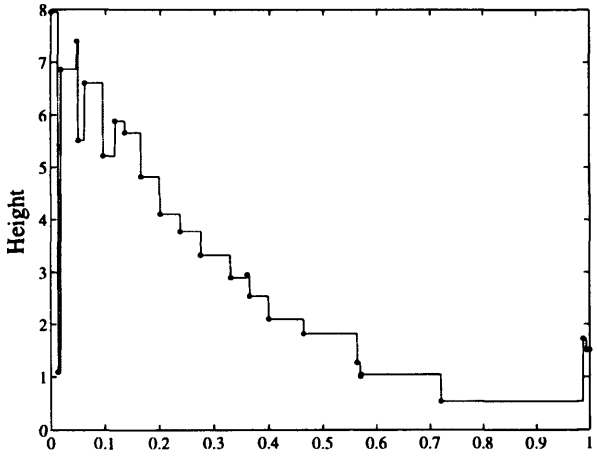


Figure 5.4: A realisation under the updating scheme of chapter (4).

5.4 Applications

In this section we investigate the performance of the method in three situations, where we use the real world dataset and the two simulated samples presented in chapter 4 for the estimation of the copula density. We remind that the two simulated data, with size 2000 and 500 observations which we discuss first, come from the same distribution,

$$C(u,v) = \left[1 + \left((u^{-0.5} - 1)^{1.5} + (v^{-0.5} - 1)^{1.5}\right)^{1/1.5}\right]^{-1/0.5}. \qquad (5.8)$$

The real world dataset concerns the daily readings of the DowJones and FTSE share indices. For reasons of convenience, we include again the scatter plots of the generated observations in figure (5.5).

For both datasets, we run the sampler long enough to assume that convergence has occurred. The hyperparameter values in all cases are the same, $\mu = 25, \gamma = 0.4, C_\varepsilon = 25, \delta = 0.08$ and $\tau = 8$. We start collecting an MCMC sample of $M = 1000$ realisations after an initial run of the chain for a burn-in

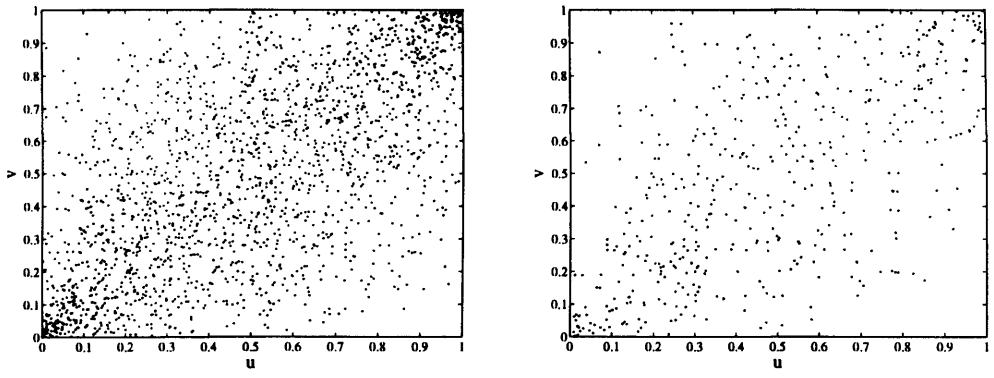


Figure 5.5: Two simulated datasets. The left plot has size 2000 observations and the right 500.

period of 200000 iterations. The Monte Carlo simulation was run for a total of 1200000 updates and each realisation was collected after every 1000 iterations to finally obtain the MCMC sample of quadratic splines $\phi^{(1)}, \phi^{(2)}, \ldots, \phi^{(M)}$ and the associated first and second derivatives. For the large sample, $n = 2000$ observations, the simulation, developed in MATLAB 6.5, took 12 hours to complete on an computer with CPU an AthlonXP 1800 and 768MB RAM under WindowsXP environment.

Bayesian techniques focus on the whole posterior distribution, and not just a point. The MCMC sample is used here to approximate the posterior. For a point estimate of the theoretical generator we take the estimated posterior mean

$$\hat{\phi}(x) = \frac{\sum_{m=1}^{M} \phi^{(m)}(x)}{m}$$

of the quadratic generators, evaluated on 100 equidistant points in $[x_1, x_{K+1}]$. A similar summary statistic is used for the estimation of the derivatives,

$$\hat{\phi}'(x) = \frac{\sum_{m=1}^{M} \phi'^{(m)}(x)}{m} \qquad \text{and} \qquad \hat{\phi}''(x) = \frac{\sum_{m=1}^{M} \phi''^{(m)}(x)}{m}.$$

Because Archimedean copulas are invariant to scale transformations of the generator, to visually demonstrate the goodness of fit of the estimators we use the ratio $\frac{\hat{\phi}(x)}{\hat{\phi}'(x)}$ which is scalar-free and plotted in figure (5.6) against the theoretical counterpart. Figure (5.7) shows three realisations, $\phi''^{(1)}, \phi''^{(500)}$ and $\phi''^{(1000)}$, together with a graph of the theoretical second derivative. We also include in

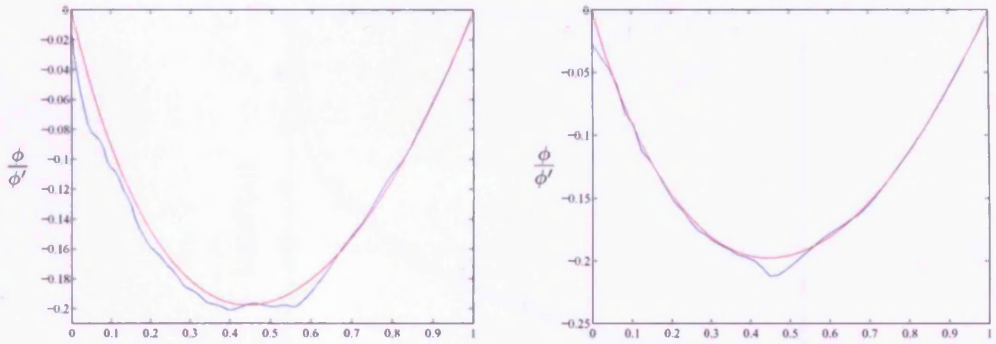


Figure 5.6: Fit of the estimated ϕ/ϕ' for the two simulated datasets. —, theoretical; —, estimated posterior mean. The left plot corresponds to sample size $n = 2000$ and the right to $n = 500$.

figure (5.8), plots of the estimated generator and its derivatives overimposed by the theoretical corresponding functions. The last two figures are in the logarithmic scale and we have removed the effect of the scaling constant to make the figures more informative. All graphs show a quite satisfactory performance of our method apart maybe for very small values. A possible explanation for the lack of fit in these cases may be due to the limiting behaviour of the theoretical generator which goes to infinity as we approach zero. Furthermore, proposals for the heights or positions of steps located close to zero is likely to give a realisation which is not convex or decreasing and therefore is rejected. This explains the existence of two very similar steps in the MCMC sample as we can see from the left part of figure (5.7). Otherwise the method seems to adapt well as flat areas are usually represented by steps defined on larger subintervals than those corresponding to steeper slopes. From the posterior mean generator and its associated derivatives we derive the posterior mean distribution shown in figure (5.9). It is apparent that the surface is smooth and does not exhibit the sudden changes that other non-parametric techiques such as the empirical estimator, produce on the estimates. We should underline though, that the scope of this chapter is mainly to propose a non-parametric method that gives smooth estimates of a convex function and not just to estimate the distribution of a bivariate dataset.

Interesting inferences for the variables under study can be drawn from the MCMC sample. For example, it is straightforward to estimate the posterior

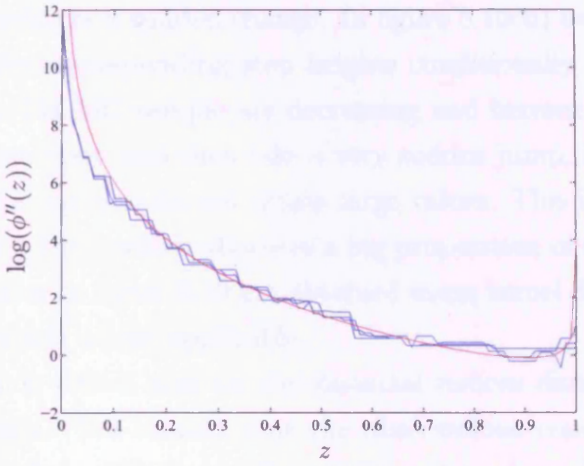
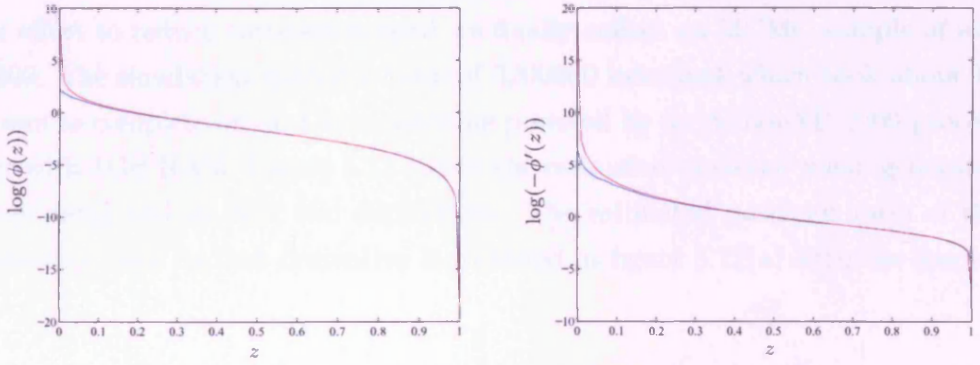


Figure 5.7: Three realisations from the posterior sample (in blue colour) overimposed by the theoretical second derivative of the copula generator (in red colour).



(a) Estimated posterior mean generator (in blue colour) against the theoretical one (in red colour).

(b) Estimated posterior mean first derivative of the generator (in blue colour) against the theoretical one (in red colour).

Figure 5.8: Estimated posterior expected generator and its derivatives. All plots are in logarithic scale.

distribution of the number of steps, $K \,|\, \text{data}$. The estimated posterior density, $p(K \,|\, \text{data})$ is shown in figure 5.10(a). Figure 5.10(b) shows the estimated density of the position of the change points conditionally on the number of the steps $K = 32, 40, 48$. Note that as the number of the steps increases, mass concentrates on the right endpoint of the domain (close to one) where the theo-

retical function presents a sudden change. In figure 5.10(b) we plot the density of logarithm of the corresponding step heights conditionally on K. Since the realisations in the MCMC sample are decreasing and become flatter as we approach the right endpoint and then take a very sudden jump, it is likely to have a high number of steps that do not attain large values. This is depicted by the estimate in figure 5.10(c) which allocates a big proposition of the mass to small values. All estimates in figure 5.10 are obtained using kernel density estimation with reflective bounds where applicable.

Our approach is tested now on the financial indices dataset already used in chapters 3 and 4. We remind that the observations concern the percentage daily returns of the FTSE and Dow-Jones indices from January 1st, 1987 until May 31st 2001. The Monte Carlo simulation was run with the same hyperparameter values as the ones used for the simulated dataset examples, $(\mu = 25, \gamma = 0.4, C_\varepsilon = 25, \delta = 0.08, \tau = 8)$. Following an initial burn-in period of 1700000 iterations, we start keeping each realisation every after 1500 runs in an effort to reduce correlation until we finally collect an MCMC sample of size 1000. The simulation needed a total of 3200000 iterations which took about 18 hours to complete on an Linux machine powered by an AthlonXP 2200 processor with 1GB RAM. Figure 5.11 shows the estimated posterior mean generator $\mathbb{E}(\phi \,|\, \text{data})$ and its first two derivatives. The estimated posterior ratio of the generator over its first derivative is included in figure 5.12(a) with the associ-

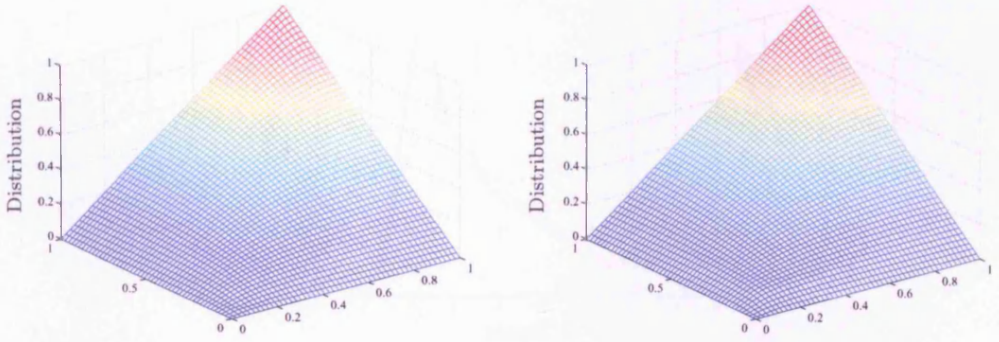
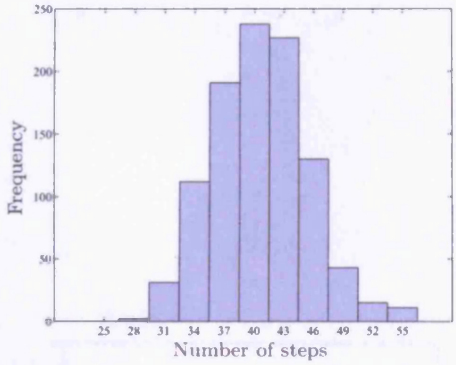
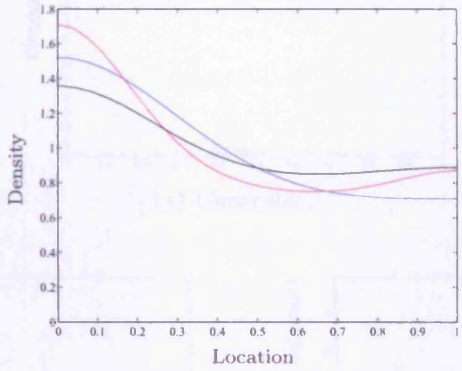


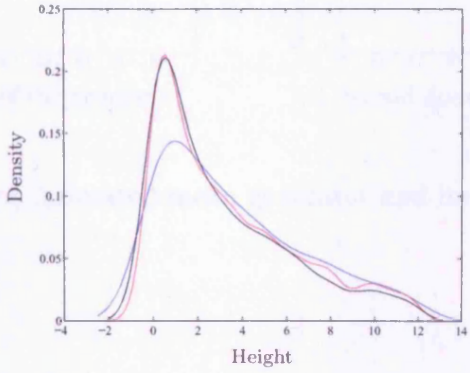
Figure 5.9: Estimated posterior expected distribution for $n = 2000$ (left) and $n = 500$ (right).



(a) Estimated posterior density of the number of steps K.

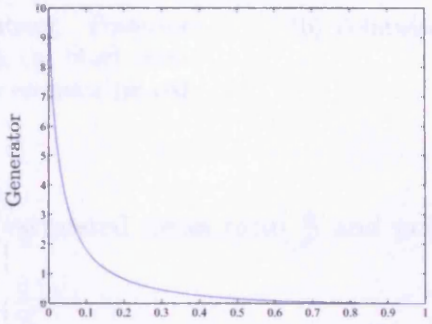


(b) Estimated posterior density of the position of the steps given $K = 32$ (blue curve), 40 (red curve), and 48 (black curve).

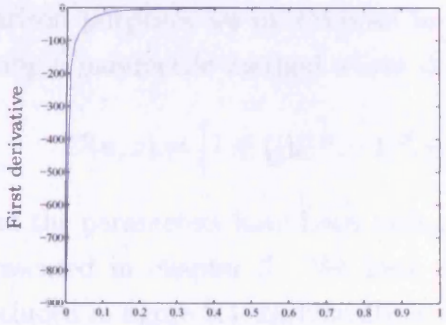


(c) Estimated posterior density of the heights (in the log scale) of the steps given $K = 32$ (blue curve), 40 (red curve), and 48 (black curve).

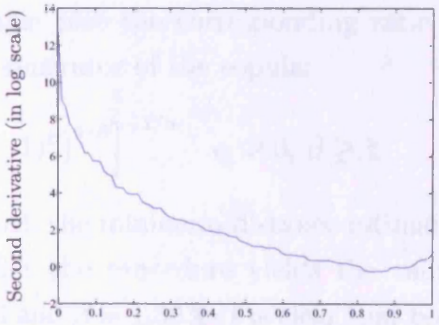
Figure 5.10: Posterior densities of some variables under interest.



(a) Generator.

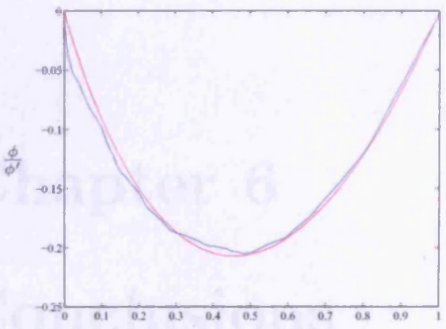


(b) First derivative of the generator.

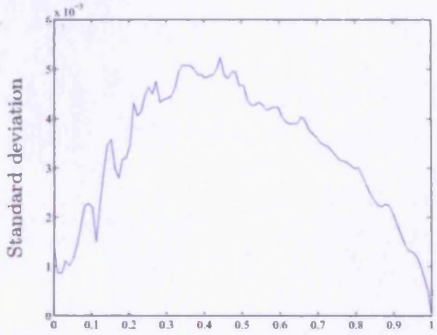


(c) Second derivative of the generator

Figure 5.11: Posterior estimated mean generator and its first two derivatives.



(a) Financial indices dataset: Posterior estimated mean ratio $\frac{\phi}{\phi'}$ (in blue) overimposed by a parametric estimate (in red colour).



(b) Pointwise standard deviations.

Figure 5.12: Posterior estimated mean ratio $\frac{\phi}{\phi'}$ and pointwise standard deviations.

ated piecewise standard deviations in the sample $\frac{\phi^{(m)}}{\phi'^{(m)}}$ plotted in figure 5.12(b). All functions are evaluated on 100 equdistant points from $[x_1, x_{K+1}]$. For comparison purposes we overimpose in the same plot the corresponding ratio $\frac{\phi}{\phi'}$ using a parametric method where ϕ is the generator of the copula:

$$C(u,v) = \left[1 + \left((u^{-\alpha}-1)^{\beta} + (v^{-\alpha}-1)^{\beta}\right)^{1/\beta}\right]^{-1/\alpha} \qquad \alpha > 0,\ \beta \geq 1$$

and the parameters have been estimated with the minimum distance estimator presented in chapter 3. We have seen that the procedure yields the values included in figure 3.4, specifically $\alpha = 0.593$ and $\beta = 1.383$. It is clear that both methods give similar results; what is encouraging though, is the closeness of the nonparametric estimate to the parametric one, something which may be due to the large sample size. However, results from smaller sample experiments such as those included in figures 4.9 and 5.6 and concern $n = 500$ bivariate simulated observations from the same copula, makes us confident that the approach we developed in this chapter is robust and its perfomance will not considerably deteriorate as n decreases.

Chapter 6

Conclusions

We have presented three approaches for the estimation of the copula of a bivariate sample. In chapter 3, the procedure is independent on the marginal distributions but assumes that the copula belongs to the parametric family of the Archimedean class. The focus is to present a method for the estimation of a vector-parameter family where the approach introduced by Genest and Rivest (1993) appears to be problematic. Simulation-based results indicate that our estimator does not lack in the one-parameter case and gives better results than its competitors, in terms of MSE, for two-parameter families. More generally, our method can be directly extended to handle multi-parameter copulas and intended future work includes a more general investigation of the estimator's performance and its theoretical asymptotic properties. In chapters 4 and 5 we relax the parametric assumption and present Bayesian techniques to estimate the joint density and the generator of an Archimedean copula respectively. The algorithm described in chapter 4 is not restricted only on the Archimedean family but can be used for the estimation of any copula in general. Prior knowledge of this generic family is incorporated into the algorithm discussed in chapter 5. Both Bayesian methods were tested on the same datasets, one large of size $n = 2000$ and a smaller one with $n = 500$. Furthermore all three approaches, frequentistic and Bayesian, are illustrated on a real-life dataset consisting of 3760 daily observations of the values of two financial indices.

A quantitative comparison of the Bayesian methodologies is available from the cross-classification tables in figures 6.1 for $n = 2000$, and 6.2 for $n = 500$. For the sample of size 2000, we have assumed a 10×10 grid on the domain and each cell in the tables represents the number of the observations included in the

	Upper Bounds									
Upper Bounds	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	102	40	22	13	9	5	4	3	1	1
0.2	40	46	34	26	18	14	9	6	5	2
0.3	22	34	35	29	24	20	15	11	7	3
0.4	13	26	29	30	28	24	20	15	10	5
0.5	9	18	24	28	29	27	24	19	15	7
0.6	5	14	20	24	27	28	28	25	19	10
0.7	4	9	15	20	24	28	30	30	26	14
0.8	3	6	11	15	19	25	30	35	33	23
0.9	1	5	7	10	15	19	26	33	45	39
1.0	1	2	3	5	7	10	14	23	39	96

(a) true

	Upper Bounds									
Upper Bounds	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	103	44	21	10	6	6	4	5	3	0
0.2	43	34	34	29	23	13	11	6	2	2
0.3	21	43	36	31	32	16	19	13	11	2
0.4	15	23	28	37	30	23	20	21	11	5
0.5	6	19	23	24	28	25	29	19	14	8
0.6	7	20	21	27	32	28	32	23	21	14
0.7	6	10	15	24	18	25	31	31	21	9
0.8	3	8	10	10	16	27	32	27	33	19
0.9	1	4	8	9	5	28	24	33	37	39
1.0	0	2	3	5	11	8	7	13	37	95

(b) observed

	Upper Bounds									
Upper Bounds	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	78	51	26	16	10	8	5	3	2	1
0.2	51	37	31	24	20	12	11	8	4	2
0.3	26	31	34	31	23	19	14	11	8	3
0.4	16	24	31	27	28	24	19	16	10	5
0.5	10	20	23	28	29	26	24	18	15	7
0.6	8	12	19	24	26	24	27	28	21	11
0.7	5	11	14	19	24	27	34	32	22	12
0.8	3	8	11	16	18	28	32	31	33	20
0.9	2	4	8	10	15	21	22	33	45	40
1.0	1	2	3	5	7	11	12	20	40	99

(c) posterior spline.

	Upper Bounds									
Upper Bounds	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.1	81	42	26	15	9	8	7	3	2	2
0.2	39	38	33	24	20	19	11	9	5	1
0.3	20	33	35	29	24	20	18	10	8	4
0.4	12	30	33	32	28	25	20	12	9	8
0.5	7	23	31	31	28	27	23	19	10	8
0.6	7	16	21	24	27	29	28	26	26	10
0.7	6	13	17	20	27	28	29	31	27	11
0.8	4	8	15	18	20	27	30	29	29	15
0.9	3	4	11	12	15	21	22	30	37	35
1	2	3	4	6	9	15	12	20	37	73

(d) posterior density.

Figure 6.1: Four cross-classification tables for the dataset with size $n = 2000$. The two top tables show the number of observations that the theoretical and the empirical distributions assign in each cell. The two bottom tables are derived either by means of the estimated generator or the estimated joint copula density.

corresponding cell. The entries in the top table correspond to the theoretical values whereas those in the second one are the observed counts. Tables 6.1(c) and 6.1(d) were obtained from the posterior mean generator, using the spline estimation method, chapter 5 and the posterior mean density, chapter 4 respectively. In a similar manner we obtain the tables in figure 6.2 where this time we take a 5×5 grid and the top half of the figure includes the true and the observed table whereas the two estimated ones are shown in the bottom half.

For every Archimedean copula it is true that $c(u, v) = c(v, u)$ for all pairs (u, v) and therefore the density c is symmetric with respect to the line $u = v$. This property is reflected in table 6.1(a), however the observed table, 6.1(b), shows some minor discrepancies from this symmetry despite the relatively big sample size.

The two estimated tables in the same figure show some interesting results. Both methods seem to underestimate the density on the bottom left corner of the domain. The posterior mean generator gives the lowest estimate for this specific area, but it behaves quite well on the opposite corner where both coordinates approach $(1, 1)$ and the posterior mean density still gives underestimated values. Note that the theoretical distribution assumes a non-symmetric pattern for those two quadrants and this asymmetry is captured better by the join density estimation method than the splines approach. Because of the implicit specification of the generator ϕ as a random quadratic spline, the latter estimator is bound to satisfy the symmetry around the main diagonal $u = v$. On the other hand the estimator described in chapter 4 is designed to move freely within a wider functional space. These conceptual differences become clearly apparent if we compare table 6.1(c) which completely satisfies the symmetry around the main diagonal, with table 6.1(d) which, being constraint-free, shows values in the corresponding cells that differ. As a result each estimator may work better than the other on different occasions, where for example, the Archimedean assumption is questionable.

For the small dataset, conclusions seem to differ slightly since the two quadrants $[0.0, 0.2] \times [0.0, 0.2]$ and $[0.8, 1.0] \times [0.8, 1.0]$ show observed values in table 6.2(b) that imply the reverse dependence pattern than the true one and is described in table 6.2(a). The estimated posterior mean density follows the observed pattern but gives underestimated values as table 6.2(d) reveals. On the

Upper Bounds	Upper Bounds				
	0.2	0.4	0.6	0.8	1.0
0.2	57	24	11	6	2
0.4	24	30	25	14	7
0.6	11	25	27	25	12
0.8	6	14	25	31	24
1.0	2	7	12	24	55

(a) true.

Upper Bounds	Upper Bounds				
	0.2	0.4	0.6	0.8	1.0
0.2	49	23	9	6	1
0.4	28	40	30	17	9
0.6	14	25	29	27	15
0.8	7	16	23	27	16
1.0	2	3	4	27	53

(b) observed.

Upper Bounds	Upper Bounds				
	0.2	0.4	0.6	0.8	1.0
0.2	57	23	12	6	2
0.4	23	32	23	15	7
0.6	12	23	28	25	12
0.8	6	15	25	30	24
1.0	2	7	12	24	55

(c) posterior estimated spline.

Upper Bounds	Upper Bounds				
	0.2	0.4	0.6	0.8	1.0
0.2	44	31	13	7	3
0.4	22	35	23	18	5
0.6	10	29	28	19	8
0.8	7	18	28	27	28
1.0	3	12	16	20	46

(d) posterior estimated density.

Figure 6.2: Four cross-classification tables for the dataset with size $n = 500$.

contrary, the density recovered from the posterior mean generator gives results similar to the true values within the aforementioned rectangles and in general seems to follow the theoretical and the empirical densities closer.

In table 6.3 we include some results to compare the two methods. To quantify the fit of the two estimators to the data, we use the χ^2 statistic $\sum_{i=1}^{N}(o_i - e_i)^2/e_i$ where o_i is the i-th entry of the observed tables and e_i is the corresponding value from the two estimated matrices in either figure 6.1 or 6.2 where $N = 100$ and $N = 25$ respectively. Additionally, we investigate which estimator, including the empirical, recovers the theoretical distribution more precisely, using as a measure of proximity the sum of squared errors $\sum_{i=1}^{N}(t_i - e_i)^2$, where t_i is the

		spline estimation	density estimation	empirical
n = 2000	χ^2-fit to data	**83.67**	104.82	-
	sum of squared errors	**1222.0**	1756.0	1402.0
n = 500	χ^2-fit to data	**21.30**	50.61	-
	sum of squared errors	**20.0**	544.0	432.0

Figure 6.3: Table of summary statistics.

i-th entry in the true matrices and e_i the same as before but now refers to the empirical values too. Both summary statistics for sample size $n = 2000$ or $n = 500$, are included in table 6.3. The results show that the density obtained from the estimated posterior mean generator captures the observed dataset and also restores the true ϕ better than the estimator that assumes directly a density on the data in terms of either averaging an MCMC sample as we do in chapter 4, or using the raw empirical density. Another interesting result emerges if we compare the statistics' values for the first two estimators in table 6.3 with the benchmark values of the empirical estimator. We see that the spline estimation method does not deteriorate in performance as fast as its alternative when the sample size decreases. The reason seems to be in the design. The reader may recall that in chapter 4 the proposed density differs from the current only in the area that includes a properly chosen set and its neighbours and the density everywhere else remains unchanged up to a scalar. As a result, datapoints outside this area seem to be less informative in the computation of the likelihood ratios. In chapter 5 the proposed realisation for ϕ'' is obtained by recommending changes that concern a randomly chosen step only, which might already exists or generated during the procedure. However, the corresponding proposals for ϕ and ϕ' do not differ from the current realisations only on the interval where the step under consideration is defined but on a wider set. Consequently, in the likelihood ratio computations every observed datapoint contributes. The estimator, as an inference tool, seems to gain in power because it becomes more flexible and moves within the space with bigger steps, but looses in computational efficiency when compared with its alternative from chapter 4. Moreover, the rate of convergence deteriorates when derivatives of the function are estimated, thats why in chapter 5 we adopt a very long run to simulate the chain . However, the computational burden is counterbalanced by the fact that the objects under consideration are simple one-dimensional functions as opposed with the approach that builds estimates directly on the two-dimensional copula density.

As a final comment we note that one of our investigations concerned the implementation of a fully locally updating scheme for all ϕ, ϕ', ϕ'' but convergence was very slow. It is also interesting to study the performance of the method within a wider framework as an inference device for convex functions in general

and not only for copula generators. The price of a call option for example, must be a decreasing and convex function of the options's strike price. Aït-Sahalia and Duarte (2003) use a two step procedure where they first use constrained least squares regression to obtain what they call a set of transformed data points and then construct the estimator by employing locally polynomial kernel smoothing. Our approach offers the potential for investigations for similar problems from a different Bayesian perspective.

Bibliography

Abramowitz, M., and I.A. Stegun. 1972. *Handbook of Mathematical Functions.* New York: Dover Publications, Inc.

Aït-Sahalia, Y., and J. Duarte. 2003. Nonparametric option pricing under shape restrictions. *Journal of Econometrics* 116:9–47.

Al-Awadhi, F., F. Hurn, and C. Jennison. 2001. Improving the acceptance rate of reversible jump MCMC proposals. Research Report, University of Bath.

Aurenhammer, F. 1991. Voronoi diagrams - a survey of a fundamental geometric data structure. *ACM Computing Surveys* 23:345–405.

Besag, J. 1986. On the statistical analysis of dirty pictures. *Journal of Royal Statistical Society: Series B* 48(3):259–302.

Besag, J., P. Green, D. Higdon, and K. Mengersen. 1995. Bayesian computation and stochastic systems. *Statistical Science* 10(1):3–66.

Blum, J.R., J. Kiefer, and M. Rosenblatt. 1961. Distribution-free tests of independence based on the sample distribution function. *Ann. Math. Statist.* 32: 485–498.

Brooks, S., P. Guidici, and G.O. Roberts. 2002. Efficient construction of reversible jump MCMC proposal distributions. *Journal of Royal Statistical Society: Series B* 65:1–37.

Casella, G., and R.L. Berger. 2002. *Statistical Inference.* 2nd ed. Wadsworth Group. CA.

Chou, P.B., and C.M. Brown. 1990. The theory and practice of Bayesian image modeling. *International Journal of Computer Vision* 4:185–210.

Clayton, D.G. 1978. A model for association in bivariate life tables and its application in epidemiological studies of familiar tendency in chronic disease incidence. *Biometrica* 65:141–151.

Cook, R.D., and M.E. Johnson. 1981. A family of distributions for modelling non-elliptically symmetric multivariate data. *Journal of Royal Statistical Society: Series B* 43(2):210–218.

Dall'Aglio, G. 1959. Sulla compatibilitá delle funzioni di ripartizione doppia. *Rend. Mat.* 18:385–413.

Darsow, W. F., B. Nguyen, and E.T. Olsen. 1992. Copulas and Markov processes. *Illinois Journal Of Mathematics* 36(4):601–642.

Deheuvels, P. 1978. Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes. *Publications de l'Institut de Statistique de l'Université de Paris* 23:1–36.

———. 1979. La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Bulletin de l'Académie Royale de Belgique. Classe de Sciences.* 5(65):274–292.

Drouet-Mari, D., and S. Kotz. 2001. *Correlation and Dependence.* London: Imperial College Press.

Embrechts, P., A.J. McNeil, and D. Straumann. 2002. Correlation and dependence in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond*, ed. M.A.H. Dempster, 176–223. Cambridge University Press.

Feller, W. 1968. *An Introduction to Probability Theory and its Applications*, vol. I. 3rd ed. London: Wiley.

———. 1971. *An Introduction to Probability Theory and its Applications*, vol. II. 2nd ed. London: Wiley.

Féron, R. 1956. Sur les tableaux de corrélation dont les marges sont données, cas de l'espace à trois dimensions. *Publ. Inst. Statist. Univ. Paris* 5:3–12.

Fisher, N.I. 1998. Copulas. In *Encyclopedia of Statistical Sciences*, ed. I.S. Kotz, C.B. Read, and D.L Banks, vol. Update Vol. 1, 159–163. John Wiley and Sons, New York.

Frank, M.J. 1979. On the simultaneous associativity of $f(x, y)$ and $x+y-f(x, y)$. *Aequationes Mathematicae* 19:194–226.

Fréchet, M. 1951. Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon Sect. A* 9:53–77.

Frees, E. W., and E. A. Valdez. 1998. Understanding relationships using copulas. *North American Actuarial Journal* 2(1):1–25.

Geman, D., S. Geman, and P. Dong. 1990. Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(7):609–628.

Geman, S., and D. Geman. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6:721–741.

Genest, C. 1987. Frank's family of bivariate distributions. *Biometrica* 74:549–555.

Genest, C., K. Ghoudi, and L.P. Rivest. 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82(3):543–552.

Genest, C., and J. MacKay. 1986a. The joy of copulas: Bivariate distributions with uniform marginals. *American Statistician* 40:280–283.

Genest, C., and J.R. MacKay. 1986b. Copules Archimédiennes et familles de lois bidimesionnelles dont les marges sont donneés. *The Canadian Journal of Statistics* 14:145–159.

Genest, C., and J.F. Plante. 2003. On Blest's measure of rank correlation. *The Canadian Journal of Statistics* 31:1–18.

Genest, C., and L. Rivest. 1993. Statistical inference procedures for bivariate archimedean copulas. *Journal of American Statistical Association* 88:1034–1043.

Geyer, C. 1992. Practical Markov Chain Monte Carlo. *Statistical Science* 7(4): 473–483.

Green, P., and A. Mira. 2001. Delayed rejection in reversible jump Metropolis-Hastings. Research Report, University of Bristol.

Green, P.J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82(4):711–732.

Gumbel, E.J. 1960a. Bivariate exponential distributions. *Journal of the American Statistical Distribution* 55:698–707.

———. 1960b. Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris* 9:171–173.

———. 1961. Bivariate logistic distributions. *Journal of the American Statistical Distribution* 56:335–349.

Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1):97–109.

Heikkinen, J. 1998. Curve and surface estimation using dynamic step functions. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, ed. D. Dey, P. Müller, and D. Sinha, chap. 14, 255–272. Spinger -Verlag.

Heikkinen, J., and E. Arjas. 1998. Non-parametric Bayesian estimation of a spatial Poisson intensity. *Scandinavian Journal of Statistics* 25:436–450.

Hoeffding, W. 1940. Masstabinvariante korrelationstheorie. Ph.D. thesis, Schriften des Matematischen Instituts fur Angewandte Matematik der Universitat Berlin. Reprinted in Hoeffding (1994).

———. 1994. Scale-invariant correlation theory. In *The Collected Works of Wassily Hoeffding.*, ed. N.I. Fisher and P.K. Sen, 57–107. Spinger-Verlag.

Hougaard, P. 1986a. A class of multivariate failure time distributions. *Biometrika* 73:671–678.

———. 1986b. Survival models for heterogeneous populations derived from stable distributions. *Biometrica* 73:387–396.

Hutchinson, T.P., and C.D. Lai. 1990. *Continuous Bivariate Distributions, Emphasising Applications.* Adelaide, South Australia: Rumsby Scientific Publishing.

Joe, H. 1997. *Multivariate Models and Dependence Concepts.* London: Chapman and Hall.

Kimeldorf, G., and A. Sampson. 1982. Uniform representations of bivariate distributions. *Communications in Statistics* 4:617–627.

Kotel'nikova, V.F., and E.V. Chmaladze. 1982. On computing the probability of an empirical process not crossing a curvilinear boundary. *Theory of Probability and its Applications* 27:640–648.

Kruskal, W.H. 1958. Ordinal measures of association. *Journal of the American Statistical Association* 53(284):814–861.

Lehmann, E.L. 1966. Some concepts of dependence. *The Annals of Mathematical Statistics* 37(5):1137–1153.

Ling, C.H. 1965. Representation of associative functions. *Publicationes Mathematicae, Debrecen* 12:189–212.

Liu, J.S., F. Liang, and W.H. Wong. 2001. The theory of dynamic weighting in monte carlo computation. *Journal of the American Statistical Association* 96(454):561–573.

Longin, F., and B. Solnik. 2001. Extreme correlation of international equity markets. *Journal of Finance* 56(2):649–676.

Marroquin, J.L., S. Mitter, and T. Poggio. 1987. Probabilistic solution of ill-posed problems in computer vision. *International Journal of Computer Vision* 76–89.

Marshall, A.W., and I. Olkin. 1988. Families of multivariate distributions. *American Statistical Association* 83(403):834–840.

Menger, K. 1942. Statistical metrics. *Proc. Nat. Acad. Sci. U.S.A.* 28:535–537.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21:1087–1091.

Nelsen, R. 1986. Properties of a one-parameter family of bivariate distributions with specified marginals. *Comm. Statist* 15:3277–3285.

———. 1999. *An Introduction to Copulas.* New York: Springer Verlag.

Nelsen, R., J.J. Quesada-Molina, J.R. Rodríguez-Lallena, and M. Úbeda-Flores. 2003. Kendall distribution functions. *Statistics & Probability Letters* 65:263–268.

Oakes, D. 1982. A model for association in bivariate survival data. *Journal of Royal Statistical Society: Series B* 44:414–493.

Okabe, A., B. Boots, K. Sugihara, and S.N. Chiu. 2000. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams.* 2nd ed. London: John Wiley and Sons.

Rényi, A. 1959. On measures of dependence. *Acta. Math. Acad. Sci. Hungar.* 10:441–451.

Roberts, G.O., A. Gelman, and W.R. Gilks. 1997. Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability* 7(1):110–120.

Scarsini, M. 1984. On measures of concordance. *Stochastica* 8:201–218.

Schweizer, B., and A. Sklar. 1983. *Probabilistic Metric Spaces.* New York: North - Holland.

Schweizer, B., and E. Wolff. 1981. On non-parametric measures of dependence for random variables. *Annals of Statistics* 9:879–885.

Sklar, A. 1959. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* 8:229–231.

———. 1996. Random variables, distribution functions, and copulas - a personal look backward and forward. In *Distibutions with Fixed Marginals and Related Topics.*, ed. L. Ruschendorf, B. Schweizer, and M.D. Taylor, 1–14. Institute of Mathematical Statistics, Hayward, CA.

Tibiletti, L. 1996. Beneficial changes in random variables via copulas: An application to insurance. *The Geneva Papers on Risk and Insurance Theory* 20(2):3–12.

Whitt, W. 1976. Bivariate distributions with given marginals. *Ann. Statist.* 4: 1280–1289.